311
X 452 с.с.
M134202

Дж. Хили

СТАТИСТИКА

социологические и маркетинговые исследования

6-е издание

THOMSON WADSWORTH

издательство DiaSoft

ПИТЕР

Joseph F. Healey
Christopher Newport University

STATISTICS

a tool for social research

Sixth Edition

THOMSON
WADSWORTH

Australia • Canada • Mexico • Singapore • Spain
United Kingdom • United States

каф. маркетинга

Дж. Хили

СТАТИСТИКА

социологические и маркетинговые исследования

Под общей редакцией
кандидата физико-математических наук
А. А. Руденко

6-е издание

Отд. 311
Шифр Х 452
М-134202
Инв. №
БИБЛИОТЕКА

издательство DiaSoft
Москва · Санкт-Петербург · Киев
2005

ПИТЕР®
Москва · Санкт-Петербург · Нижний Новгород · Воронеж
Ростов-на-Дону · Екатеринбург · Самара · Новосибирск
Киев · Харьков · Минск
2005

ББК 22.172+60.5
УДК 519.22:30
Х-45

Хили Дж.
Х-45 Статистика. Социологические и маркетинговые исследования. 6-е изд. / Пер. с англ. Под общей ред. к. ф.-м. н. А. А. Руденко. — Киев: ООО «ДиаСофтЮП»; СПб.: Питер, 2005. — 638 с.: ил.

ISBN 5-93772-122-6
ISBN 5-469-00813-4

[illegible]

ББК 22.172+60.5
УДК 519.22:30

[illegible] Thomson Learning (EMEA) Ltd.

[illegible]

[illegible]

© Thomson Learning, 2002
© Перевод на русский язык, оформление, ООО «ДиаСофтЮП», 2005
© Обложка, ЗАО Издательский дом «Питер», 2005

Оглавление

Предисловие **15**

Глава 1. Введение **23**

1.1. Зачем нужно изучать статистику? 24
1.2. Роль статистики в научном исследовании 23
1.3. Цели книги 29
1.4. Описательная статистика и статистические выводы 30
1.5. Дискретные и непрерывные переменные 32
1.6. Шкалы измерений 33
Чтение статистик 1: Введение 37
Заключение 39
Словарь 39
Мультимедийные ресурсы 40
Задачи 41
Введение в автоматизированные статистические пакеты и Всеобщий общественный опрос 45

Часть I. Описательная статистика **47**

Глава 2. Основы описательной статистики. Процентные показатели, пропорции и удельные показатели, таблицы, диаграммы и графики **48**

2.1. Процентные показатели и пропорции 49
Практический пример 2.1 50
2.2. Отношения и удельные показатели 52
Практический пример 2.2 53
Практический пример 2.3 55
2.3. Частотные распределения: введение 55
2.4. Частотные распределения для переменных, измеренных по номинальной и порядковой шкале 57
2.5. Частотные распределения для переменных, измеряемых по интервальной/относительной шкале 59
Практический пример 2.4 59
2.6. Построение частотных распределений для интервальных/относительных переменных: обзор 65
Чтение статистик 2: Частотные распределения и диаграммы 67
2.7. Диаграммы и графики 70
2.8. Интерпретация статистик: Использование процентных показателей, частотных распределений, диаграмм и графиков для анализа изменения степени надзора за рабочим местом 75
Заключение 79
Сводка формул 79
Словарь 80
Мультимедийные ресурсы 81
Задачи 81

Создание частотных распределений и графиков с помощью SPSS for Windows ... 88

Глава 3. Показатели центра распределения ... 96

3.1. Введение ... 96
3.2. Мода ... 97
3.3. Медиана ... 99
3.4. Прочие показатели положения: процентиль, дециль, квартиль ... 99
3.5. Среднее арифметическое ... 100
3.6. Некоторые свойства среднего арифметического ... 102
Практический пример 3.1 ... 104
3.7. Выбор показателя центра распределения ... 105
Заключение ... 106
Формулы ... 107
Словарь ... 107
Мультимедийные ресурсы ... 107
Задачи ... 108
Использование *SPSS for Windows* для вычисления центра распределения и приложение ... 115

Глава 4. Меры разброса ... 120

4.1. Введение ... 120
4.2. Коэффициент изменчивости категорий (IQV) ... 121
4.3. Размах (R) и интерквартильная широта (Q) ... 123
4.4. Расчет размаха и интерквартильной широты ... 125
4.5. Стандартное отклонение ... 126
4.6. Удобная формула для расчета стандартного отклонения ... 129
Практический пример 4.1 ... 131
Практический пример 4.2 ... 132
4.7. Интерпретация стандартного отклонения ... 132
4.8. Интерпретация статистик: использование мер центра распределения и мер разброса для анализа этнической неоднородности населения США ... 133
Чтение статистик 3: Меры центра распределения и меры разброса ... 134
Заключение ... 143
Сводка формул ... 143
Словарь ... 144
Мультимедийные ресурсы ... 144
Задачи ... 145
Использование *SPSS for Windows* для расчета мер разброса ... 148

Глава 5. Кривая нормального распределения ... 154

5.1. Введение ... 154
5.2. Расчет стандартизованных значений (Z-значений) ... 156
5.3. Таблица кривой нормального распределения ... 157
5.4. Расчет общей площади перед и после значения ... 159
5.5. Определение областей между двумя значениями ... 161
Практический пример 5.1 ... 161
5.6. Использование кривой нормального распределения для оценки вероятностей ... 163

Заключение ... 165
Формулы ... 166
Словарь ... 166
Мультимедийные ресурсы ... 166
Задачи ... 167
Использование *SPSS for Windows* для преобразования исходных значений в стандартизированные ... 170

Итоговые упражнения к части I ... 173

Часть II. Статистические выводы ... 177

Глава 6. Введение в статистические выводы. Формирование выборок и распределение выборочных статистик ... 178

6.1. Введение ... 178
6.2. Методы формирования вероятностных выборок ... 179
6.3. Методы формирования выборок по принципу МРВО ... 180
6.4. Распределение выборочных статистик ... 184
6.5. Обозначения и терминология ... 188
Заключение ... 189
Глоссарий ... 190
Мультимедийные ресурсы ... 191
Задачи ... 192
Использование *SPSS for Windows* для формирования случайных выборок ... 193

Глава 7. Процедуры оценивания ... 196

7.1. Введение ... 196
7.2. Смещение и эффективность ... 197
7.3. Процедуры оценивания: введение ... 200
7.4. Процедуры интервального оценивания выборочных средних ... 202
7.5. Процедуры интервального оценивания для выборочных пропорций (большие выборки) ... 205
Практический пример 7.1 ... 206
Чтение статистик 4: Опросы общественного мнения ... 207
7.6. Управление шириной интервальных оценок ... 208
Практический пример 7.2 ... 211
Чтение статистик 5: Отслеживание национальных тенденций с помощью репрезентативных выборок ... 211
7.7. Интерпретация статистик: Доверительные интервалы, опросы общественного мнения и президентские выборы 2000 года ... 215
Заключение ... 217
Сводка формул ... 217
Словарь ... 218
Мультимедийные ресурсы ... 218
Задачи ... 218
Построение доверительных интервалов с помощью процедур SPSS for Windows ... 220

Глава 8. Проверка гипотез I. Случай одной выборки ... 225

8.1. Введение ... 225
8.2. Общий обзор процесса проверки гипотез ... 227

8.3. Пятишаговая модель проверки гипотез ... 231
8.4. Односторонняя и двусторонняя проверка гипотез ... 234
8.5. Выбор значения доверительного интервала альфа ... 239
8.6. Распределение Стьюдента ... 241
Практический пример 8.1 ... 245
8.7. Проверка гипотез для долей одной выборки (большие выборки) ... 246
Практический пример 8.2 ... 247
Заключение ... 250
Сводка формул ... 251
Словарь ... 251
Мультимедийные ресурсы ... 252
Задачи ... 253

Глава 9. Проверка гипотез II. Случай с двух выборок ... 259

9.1. Введение ... 259
9.2. Проверка гипотез с помощью средних значений выборки (большие выборки) ... 260
9.3. Проверка гипотез с помощью средних значений выборки (небольшие выборки) ... 264
Практический пример 9.1 ... 267
9.4. Проверка гипотез с помощью долей выборки (большие выборки) ... 268
Практический пример 9.2 ... 270
Чтение статистик 6: Проверка гипотез ... 272
9.5. Ограничения в проверке гипотез: Статистическая значимость и практическая важность ... 274
9.6. Интерпретация статистик: существуют ли значимые различия между доходами мужчин и доходами женщин? ... 276
Заключение ... 279
Сводка формул ... 280
Словарь ... 281
Мультимедийные ресурсы ... 281
Задачи ... 282
Использование *SPSS for Windows* для проверки значимости различий между двумя средними значениями ... 288

Глава 10. Проверка статистических гипотез III: Дисперсионный анализ ... 293

10.1. Введение ... 293
10.2. Логика дисперсионного анализа ... 294
10.3. Вычисления при применении ДА ... 296
10.4. Упрощенная схема вычислений ... 298
10.5. Пример вычислений для ДА ... 299
10.6. Критерий значимости для ДА ... 300
Практический пример 10.1 ... 301
10.7. Дополнительный пример на применение ДА, иллюстрирующий расчет статистики критерия и применение критерия ... 304
10.8. Ограничения на применение ДА ... 307
Чтение статистик 7: Исследование отношения работодателей к инвалидам ... 308
10.9. Интерпретация статистик: меняется ли сексуальная активность в зависимости от семейного положения? ... 309

Заключение ... 313
Сводка формул ... 314
Словарь ... 315
Мультимедийные ресурсы ... 315
Задачи ... 316
Использование *SPSS for Windows* для проведения дисперсионного анализа ... 322

Глава 11. Проверка статистических гипотез IV: Критерий хи-квадрат ... 327

11.1. Введение ... 327
11.2. Двухвходовые таблицы сопряженности ... 328
11.3. Логика применения критерия хи-квадрат ... 330
11.4. Расчет статистики для критерия независимости хи-квадрат ... 331
11.5. Критерий независимости хи-квадрат ... 333
11.6. Пример применения критерия независимости хи-квадрат ... 335
Пактический пример 11.1 ... 337
Чтение статистик 8: Еще одно исследование влияния пола на доход ... 339
11.7. Ограничения на применения критерия независимости хи-квадрат ... 340
11.8. Интерпретация статистик: смертная казнь и уровень тяжелой преступности ... 342
Заключение ... 346
Сводка формул ... 346
Словарь ... 347
Мультимедийные ресурсы ... 348
Задачи ... 348
Использование *SPSS for Windows*: критерий независимости хи-квадрат ... 354
Итоговые упражнения к части II ... 359

Часть III. Меры связи ... 363

Глава 12. Связь между двумя переменными. Введение и основные понятия ... 364

12.1. Статистическая значимость и теоретическая важность ... 364
12.2. Связь между переменными и таблица сопряженности ... 365
12.3. Три характеристики парных статистических связей ... 367
Практический пример 12.1 ... 370
Чтение статистик 9: Таблицы сопряженности ... 371
Заключение ... 375
Словарь ... 376
Мультимедийные ресурсы ... 377
Задачи ... 377
Использование *SPSS for Windows* для анализа связей между двумя переменными ... 383

Глава 13. Связь между переменными, измеряемыми по номинальной шкале ... 390

13.1. Введение ... 390
13.2. Меры связи на основе критерия хи-квадрат ... 391
Чтение статистик 10: Важность процентных величин ... 394
13.3. Пропорциональное уменьшение ошибки ... 396
13.4. Меры пропорционального уменьшения ошибки для номинальных переменных: лямбда ... 397

.... 399

13.5. Вычисление лямбды 400
Практический пример 13.1 401
13.6. Ограничения применения критерия лямбда 402
Заключение 402
Сводка формул 402
Словарь 403
Мультимедийные ресурсы 403
Задачи
Использование *SPSS for Windows* для вычисления мер связи между номинальными переменными 410

Глава 14. Связь между переменными, измеряемыми по порядковой шкале **414**
14.1. Введение 414
14.2. Пропорциональное уменьшение ошибки (PRE) 416
14.3. Вычисление коэффициента гамма 416
Практический пример 14.1 420
Чтение статистик 11: Двумерные таблицы сопряженности и соответствующие статистики 421
14.4. Определение направления связи 422
14.5. Интерпретация статистической связи с помощью таблиц сопряженности на примере исследования участия граждан США в общественной жизни 426
14.6. Коэффициент ро Спирмена (rs) 429
Практический пример 14.2 433
14.7. Проверка нулевой гипотезы об отсутствии связи посредством коэффициентов гамма и ро Спирмена 434
Заключение 437
Сводка формул 437
Словарь 437
Мультимедийные ресурсы 437
Задачи 438
Использование *SPSS for Windows* для вычисления порядковых мер связи 445

Глава 15. Связь между переменными, измеренными по интервальной или относительной шкале **450**
15.1. Введение 450
15.2. Диаграммы рассеяния 451
15.3. Регрессия и предсказание 454
15.4. Вычисления *a* и *b* 457
15.5. Коэффициент корреляции (*r* Пирсона) 460
15.6. Интерпретация коэффициента корреляции: коэффициент детерминации r^2 461
Практический пример 15.1 463
15.7. Проверка коэффициента корреляции r Пирсона на значимость 465
15.8. Интерпретация статистик: действительно ли в экономике США господствуют тесно взаимосвязанные корпорации? 467
Заключение 470
Сводка формул 471
Словарь 472

Мультимедийные ресурсы 472
Задачи 473
Использование *SPSS for Windows* для вычисления r Пирсона 475
Итоговые упражнения к части III 483

Часть IV. Многомерные методы **487**

Глава 16. Детальная разработка таблиц сопряженности **488**
16.1. Введение 488
16.2. Управление с помощью третьей переменной 489
16.3. Интерпретация частных таблиц сопряженности 492
Практический пример 16.1 494
16.4. Частный показатель коэффициент гамма (Gp) 500
16.5. Откуда берутся управляющие переменные? 502
16.6. Ограничения при детальной разработке таблиц сопряженности 504
16.7. Интерпретация статистик: анализ готовности принять участие в политических акциях 505
Заключение 507
Сводка формул 508
Словарь 508
Мультимедийные ресурсы 509
Задачи 509
Использование *SPSS for Windows* для детальной разработки таблиц сопряженности 521

Глава 17. Частная корреляция, множественная регрессия и корреляция **524**
17.1. Введение 524
17.2. Частная корреляция 525
17.3. Множественная регрессия: предсказание значений зависимой переменной 528
17.4. Множественная регрессия: оценка влияния независимых переменных 532
Чтение статистик 12: Регрессия и корреляция 534
17.5. Множественная корреляция 536
17.6. Интерпретация статистик: действительно ли размер компании имеет значение для ассоциированного управления? 538
Практический пример 17.1 539
17.7. Ограничение применимости методов множественной регрессии и корреляции 544
Чтение статистик 13: Причины подростковой преступности 545
Заключение 546
Сводка формул 547
Словарь 549
Мультимедийные ресурсы 549
Задачи 549
Использование *SPSS for Windows* для регрессионного анализа 555
Итоговые упражнения к части IV 559

Приложение А. Площадь под кривой нормального распределения 561

Приложение Б. t-распределение 567

Приложение В. Распределение хи-квадрат 569
Приложение Г. F-распределение 571
Приложение Д. Использование статистик: идеи для исследовательских проектов 574
Приложение Е. Введение в программы статистического анализа для ПК (*SPSS for Windows* и *MicroCase*) 582
Приложение Ж. Кодировочная таблица для GSS 1998 594
Приложение З. Сводка основных статистических формул и условных обозначений ..... [illegible]
Ответы на расчетные задачи с нечетными номерами 608
Глоссарий терминов 625
Предметный указатель 635

Предисловие к русскому переводу

Книга Джозефа Хили является в некотором смысле уникальной и своевременной. В последнее время социология и другие общественные науки, как инструмент изучения общества, становится все более востребованными. Так, например, в демократических обществах посредством социологии осуществляется "обратная связь" между властью и народом. Она также широко используется в качестве измерительного инструмента симпатий избирателей во время выборов. Однако на современном этапе развития науки добыть необходимую информацию и принять на основе ее научно обоснованное решение можно лишь используя современные статистические подходы и методы.

Данная книга является учебником по применению статистических методов в социологических исследованиях. Учебный материал подается в нем структурированно и таким образом, что его можно изучить самостоятельно. Обилие примеров из разных областей применения социологических знаний позволяет сформировать правильное представление о возможностях применения данных методов.

Стиль написания книги таков, что даже математически неподготовленный читатель может постигнуть суть излагаемых статистических методов и подходов. Это особенно важно для читателей, работающих в области гуманитарных наук, одной из каких и является социология.

Большое внимание в данной книге автор уделяет обработке так называемых качественных или нечисловых данных. Это очень важно, так как при проведении социологических исследований основная масса данных и является таковыми, поэтому приведенные методы для их обработки являются особенно полезными.

К сожалению, во введении к 1 главе автор, говоря об ограниченности роли статистики при проведении социологических исследований, пишет, что она нужна лишь в конце этапа наблюдений. Это не совсем правильно, так как на этапе планирования исследования необходимо также применять методы теории выборочных исследований и планирования эксперимента для обеспечения репрезентативности получаемой выборки. Здесь автор имеет ввиду лишь обработку результатов выборочных исследований. Однако не следует забывать, что получение научно обоснованных результатов на основании проводимых исследований невозможно без надлежащего их планирования. Кроме того, необходимо помнить, что при постановке задачи в предметной области, в данном случае в социологии, следует учитывать ограниченность статистических методов, так как без правильно поставленного вопроса нельзя получить правильный ответ. И если вопрос исследования сформулирован без учета возможностей статистических методов, которые планируется применить для обработки результатов исследования, то в результате исследователь может и не получить корректные результаты и лишь выбросить на ветер время и деньги, потраченные на проведение исследования.

В некоторой степени непривычной для нашего читателя является классификация автором статистических методов. Так автор в подразделе 1.4 говорит о методах описательной статистики. Обычно под таковыми подразумеваются показатели, описывающие параметры распределения исследуемой переменной — форму, положение центра распределения, его асимметрию и т.д. Автор же относит к таковым и также

методы анализа данных, как корреляционный, регрессионный и другие виды анализа, цель которых состоит в выявлении и измерении зависимостей и связей между переменными.

Но все вышесказанное не умаляет достоинства данной книги как практического пособия по применению статистических методов, а лишь демонстрирует взгляд автора на некоторые дискуссионные вопросы.

Наличие в данной книге описания применения излагаемых методов при помощи современных статистических пакетов (SPSS и MicroCase) имеет важное значение. Следует отметить, что перевод приводимых в книге результатов применения этих пакетов является авторским и может не соответствовать имеющимся локализованным версиям этих пакетов. Кроме того, так как статистический пакет MicroCase в странах СНГ применяется довольно редко, то материалы, связанные с его применением, было решено представить на сайте поддержки данной книжки по адресу: www.[illegible].

В заключение следует отметить, что хотя данная книга и ориентирована на курсы социологии, читаемые в Западных высших учебных заведениях, однако она будет очень полезна для преподавателей и студентов высших учебных заведений стран СНГ, а также для социологов-практиков.

Предисловие

Социология и другие общественные науки (включая политологию, общественную деятельность, государственное управление, уголовное право, исследования народонаселения и геронтологию) являются дисциплинами, основанными на исследовании данных, и статистика является частью их повседневного языка. Чтобы освоить этот язык, вы должны быть хорошо подготовлены по методологии проведения исследований, анализа данных и подходам, применяемым в научном мышлении. Свободное владение статистикой поможет вам в понимании исследовательских отчетов, с которыми вы будете сталкиваться каждый день, и профессиональной исследовательской литературы по вашей конкретной дисциплине. Вы также сможете провести количественные исследования, внести свой вклад в развитие области социальных наук и полностью реализовать себя в качестве специалиста в сфере общественных наук.

Хотя статистика является одной из базовых дисциплин, ее изучение (и преподавание) может оказаться сложной задачей. Студенты, слушающие курс статистики, обычно имеют различный уровень знаний по математике и очень разнообразные цели в плане будущей карьеры. Часто они бывают озадачены тем, насколько важна для них статистика, и нередко они испытывают беспокойство по поводу недостаточного знания математики.

Материал данной книги направлен на решение этой сложной задачи: он содержит принципы использования статистического анализа в сфере общественных наук. Данная книга требует минимальных предварительных знаний по математике (способность прочесть простейшую формулу является единственным условием для понимания всего того, что описано в этой книге), а разнообразие дополнительных разделов помогут студентам успешно провести анализ данных. Книга написана специально для использования в программах по изучению социологии и общественных наук, но ее материал является довольно гибким и может быть использован в любой образовательной программе, относящейся к общественным наукам.

Материал изложен на уровне между строгим математическим подходом и книгой кулинарных рецептов. Я не пожертвовал всесторонним охватом дисциплины или статистической точностью, но теоретические и математические объяснения были сведены к элементарному уровню, как это и должно быть на начальной стадии изучения социальной статистики. Например, я не рассматриваю суть формальной теории вероятностей. Вместо этого, в главах 5 и 6 я неформальным и интуитивным образом знакомлю читателя со знаниями, необходимыми для понимания статистических выводов. В книге отсутствуют утверждения, что статистика — это весело, или что материал можно освоить без особого труда. В то же время, студенты не будут по-

регружены абстрактными доказательствами, выводами формул и математическими теориями, которые на данном этапе только лишь мешают образовательному процессу.

Цель книги и изменения в данном издании

Цель этой книги заключается в том, чтобы сформировать основы статистической грамотности. Человек, ориентирующийся в статистике, понимает и высоко ценит роль статистики в исследовательском процессе, он способен производить основные вычисления, а также читать и хорошо разбираться в профессиональной исследовательской литературе своей отрасли, а также без труда воспринимать любые исследовательские отчеты, с которыми он будет сталкиваться в повседневной жизни. Эта цель осталась неизменной с момента выпуска первого издания данной книги. Однако в знак признания того факта, что простые вычисления в наше высокотехнологичное время уже не представляет собой настолько сложную задачу, в данном издании больше внимания уделяется объяснению теоретического материала и компьютерным приложениям, и снят акцент с вычислений. Эти изменения в некоторых случаях будут явными. Например, во многие главы был добавлен новый подраздел под названием "Интерпретация статистик". Такие подразделы не концентрируются на проведении чистых вычислений, вместе с тем, в них содержится примерно половина материала каждой главы, и они представляют подробные примеры того, что можно сказать после получения статистических данных. Для того чтобы передать смысл и развить представления читателя, в них используются реальные данные и действительные исследовательские ситуации. Кроме того, они иллюстрируют использование статистики для поиска ответов на важные вопросы. К обсуждаемым в данных разделах темам относятся: наблюдения в рабочей среде разницы доходов между мужчинами и женщинами и предупредительный эффект смертной казни.

Кроме того, в знак признания того факта, что современные технологии сделали бессмысленным проведение долгих, утомительных вычислений вручную, наборы задач в конце каждой главы были упрощены (т.е. в них применяются наборы данных меньшего размера). Более сложные задачи (т.е. те, в которых используются наборы данных большего размера) могут быть использованы читателем, однако они перенесены на web-сайт данной книги. Кроме того, впервые данное издание содержит описание работы с MicroCase, пожалуй, самым дружественным по отношению к пользователю статистическим пакетом на сегодняшнем рынке который, наряду с SPSS, является ведущим статистическим программным обеспечением для проведения исследований в сфере общественных наук. Преподаватели могут выбрать для своего курса какой-либо из этих статистических пакетов. Набор данных, использованный в компьютерных упражнениях, был приведен в соответствие с Всеобщим общественным опросом 1998 года (основной социологический опрос, проводимый в США — *прим. научн. ред.*), а дополнительные наборы данных и упражнения размещены на web-сайте.

Еще одним значительным изменением в настоящем издании является то, что для этой книги был специально создан web-сайт (*www.sociology.wadsworth.com*). На этом web-сайте, помимо прочего, представлены дополнительные задачи, упражнения с использованием пакетов SPSS и MicroCase Explorit, дополнительные наборы данных, функциональные схемы, предназначенные для того, чтобы помочь студентам выбирать соответствующие статистические процедуры для различных ситуаций, вспомогательный математический обзор (в предыдущих изданиях — приложение Н) и множество дополнительных материалов.

Хотя изменения в данном издании и значительны, их фокусом продолжает быть усвоение основ статистической грамотности. Три нижеперечисленных аспекта этой цели являются основой для обсуждения дополнительных тем в данной книге.

1. **Понимание важности статистики.** Статистически грамотный человек понимает значимость статистики для социальных исследований, может анализировать и интерпретировать значение любого статистического критерия, а также выбирать нужный статистический показатель для конкретной исследовательской цели и данного набора данных. Этот учебник прививает эти качества (в пределах ограничений, накладываемых ознакомительным характером курса) по следующим направлениям:

   - *Значимость статистики.* Глава 1 включает обсуждение роли статистики в социальных исследованиях и делает акцент на ее полезности в качестве средства анализа и управления данными, а также поиска ответа на главные вопросы исследования. Каждая задача-пример рассматривается в контексте исследовательской ситуации. Сначала ставится вопрос, а затем с помощью статистики находится ответ. Таким образом, по всему тексту делается акцент на значимости статистики для поиска ответов на вопросы. Центральная тема полезности далее подкрепляется серией блоков, озаглавленных как Приложения, каждый из которых представляет определенный способ использования статистики для ответа на вопросы.

     Почти для каждой задачи в конце главы стоит пометка, указывающая на то, к какой дисциплине сферы общественных наук она относится: **SOC** — социология, **SW** — общественная деятельность, **PS** — политология, **CJ** — криминология, **PA** — государственное управление, **GER** — геронтология. Относя задачи к определенным дисциплинам, студентам проще будет увидеть значимость статистики для своих академических интересов. Кроме того они смогут увидеть, что все эти дисциплины имеют много общих элементов с точки зрения применения статистики.

   - *Интерпретация статистик.* Для большинства студентов интерпретация (понимание того, что означает тот или иной статистический показатель) является сложной задачей. Способность интерпретировать статистические показатели можно развить только на конкретных примерах и на опыте. При рассмотрении задач-примеров я был особенно внимателен в выражении значения статистического показателя в терминах исходной задачи исс[illegible]. Чтобы усовершенствовать практические навыки читателя, в конце [illegible] главы были включены задачи, которые почти всегда предусматривают [illegible] вычисленного статистического показателя. Чтобы дать больше примеров, многие ответы, приведенные в конце книги, выражены не только [illegible] словах. Новый подраздел каждой главы "Интерпретация статистик" представ[illegible]

[illegible] Инв. № 134202 БИБЛИОТЕКА

ет дополнительные подробные примеры того, как понимать смысл статистического показателя.

- *Использование статистики: идеи для исследовательских проектов.* Приложение E — еще один новый раздел этого издания — предлагает идеи для проектов по анализу данных для самостоятельной работы студентов. Проекты требуют от студентов использования программного статистического пакета (SPSS или MicroCase) для анализа набора данных. Эти проекты можно выдавать студентам либо через определенные интервалы на протяжении семестра, либо в конце курса. Каждый проект даст студентам возможность применить на практике свои навыки в области статистики и, прежде всего, развить способность понимать и интерпретировать значение рассчитанных ими статистических показателей. Дополнительные идеи для исследовательских проектов можно найти в дополнении к данной книге *SPSS and MicroCase Companion for Statistics: A Tool for Social Research*, Sixth Edition, которое можно заказать отдельно.

2. **Владение техникой вычислений.** Из базового курса статистики студенты должны вынести способность проводить элементарный анализ данных — выполнять последовательность вычислений и приходить к правильному результату. Конечно же, компьютеры и калькуляторы в значительной мере сняли проблему вычислений. Тем не менее, вычисления неотъемлемы от статистики, и так как студенты, специализирующиеся в общественных науках, часто не имеют достаточных навыков в области вычислений, я включил несколько соответствующих разделов с целью помочь студентам справиться с этими задачами:

- *Пошаговые вычислительные алгоритмы* предусмотрены для каждого статистического показателя.
- *Достаточно большой набор задач* предусмотрен в конце каждой главы. В большинстве своем эти задачи используют произвольные вымышленные данные, которые подобраны таким образом, чтобы облегчить вычисления.
- *Итоговые упражнения* приводятся в конце каждой части для получения практических навыков в выборе, вычислении и анализе статистических показателей. В этих упражнениях представлены только наборы данных и вопросы исследования. Студенты должны самостоятельно осуществить выбор подходящих статистических показателей для данных упражнений. Многие из этих задач, как и приведенные в конце каждой главы, в данном издании были упрощены для простоты вычислений.
- *Решения* вычислительных задач с нечетными номерами приводятся отдельно, так что студенты могут сверить с ними свои ответы.
- Это издание, как и предыдущее, предполагает использование *SPSS for Windows*, что дает студентам возможность полностью воспользоваться вычислительной мощью компьютера. Однако был добавлен дополнительный статистический пакет под названием MicroCase (использующий MicroCase Explorer) с целью повысить вариативность материала. Преподаватели могут выбрать любой пакет. Эта возможность более подробно описана ниже.

3. **Способность читать профессиональную литературу по общественным наукам.** Статистически грамотный человек способен понимать и очень хорошо разбираться в исследовательских отчетах, написанных другими. Развитие этого качества на ознакомительном уровне представляет проблему, так как: (1) словарь профессиональных исследователей гораздо более лаконичен, чем язык учебника; (2) статистика, встречающаяся в такой литературе, гораздо сложнее, чем рассматриваемая на ознакомительном уровне. Чтобы ликвидировать этот разрыв, я включил в материал серию блоков, озаглавленных "Чтение статистик", которые появляются уже в главе 1. В каждом блоке я кратко описываю стиль, обычно применяемый для данного статистического показателя в отчетах, и стараюсь предупредить студентов о том, чего им следует ожидать при обращении к профессиональной литературе. Эти вставки, многие из которых были обновлены в данном издании, включают выдержки из исследовательской литературы и иллюстрируют, как статистика применяется и интерпретируется социологами.

Дополнительные возможности

Множество других возможностей делают материал более осмысленным для студентов и более полезным для преподавателей:

- *Читабельность и доступность.* Книга написана в доступном для студентов неформальном стиле, не игнорируя традиционную терминологию статистики. Задачи и примеры были составлены с целью добиться максимального интереса студентов и сфокусировать их на вопросах важности и значимости статистики. При рассмотрении более сложного материала (например, проверки гипотез) студенты сначала проходят через решение задачи-примера, прежде чем столкнуться с формальной терминологией и понятиями. В конце каждой главы подводятся итоги по поводу основных идей и формул, а также приводится словарь важных понятий. Для быстрой справки можно пользоваться списком часто используемых формул и словарем символов, находящимися в приложении I.
- *Организация и охват материала.* Материал разделен на четыре части, причем большая его часть посвящена одномерной описательной статистике, статистическим выводам и измерениям парной корреляции (степени связи). Различие между описательной статистикой и статистическими выводами проводится в первой главе и отслеживается на протяжении всей книги. При выборе статистик, включаемых в рассмотрение, я стремился обеспечить баланс между темами, основными понятиями, с которыми студенты должны быть ознакомлены, и количеством материала, который студенты могут освоить за первый (и, возможно, единственный) курс статистики, помня о том, что разные преподаватели будут акцентировать внимание на различных аспектах темы, [illegible]. Таким образом, материал охватывает весь диапазон обычной статистики, причем каждая глава разбита на разделы, так что преподаватели могут выбрать тот материал, который они хотят включить. В данном издании материал был урезан и упрощен, что было достигнуто переносом описания некоторых редко используемых методов и статистических процедур на веб-сайт книги.

- [illegible] жат в качестве путеводителей по курсу и предназначены [illegible] на наиболее важном [illegible]
- *Обзор математических навыков.* Всесторонний обзор [illegible] требуется при изуче[illegible] [illegible] этот обзор в начале курса [illegible] или обращаться к нему по мере надобности. Есть [illegible] инструмент самоконтроля, так что студенты могут проверить свой уровень [illegible]
- *Статистические методики и задачи в конце каждой главы тесно связаны.* После рассмотрения какой-либо методики студентам предлагаются характерные [illegible] задачи для практики и закрепления материала. Тут же приводятся все [illegible]
- [illegible] дачи с меньшими наборами данных представлены первыми. Часто первые [illegible] сколько задач сопровождаются полными инструкциями и советами. П[illegible] ющие задачи становятся более сложными и требуют, чтобы выбор решения (те[illegible] выбор наиболее подходящего статистического показателя [illegible] туации) все в большей степени производился студентом самостоятельно. Таким образом, каждый набор задач развивает способности р[illegible] наборы данных наибольшего размера) задачи были перенесены на web-сайт.
- *Компьютерные приложения.* Чтобы помочь студентам воспользоваться преимуществами компьютеров, данный материал параллельно описывает два различных статистических пакета в отдельных разделах в конце каждой главы. Преподаватели могут выбрать между SPSS и MicroCase. В приложении F представлено введение в статистические пакеты в целом и в SPSS и MicroCase в частности. Демонстрационные примеры в конце глав объясняют, как [illegible] зовать статистический пакет для нахождения статистического показателя, рассмотренного в главе. Также включены и упражнения для студентов по анализу данных с использованием SPSS или MicroCase. На web-сайте данной книги также доступны дополнительные упражнения с другими наборами данных. Студенческая версия SPSS распространяется в качестве дополнения к книге, а студенческая версия MicroCase доступна на сайте.
- *Реалистичные современные данные.* База данных для компьютерных приложе[illegible] [illegible] общественного опроса 1998 года. Эта база данных дает студентам возможность [illegible]

- [illegible] [illegible] материал и некоторые менее часто используемые методы, обзор основ математики, дополнительные задачи, наборы данных, таблицы случайных чисел [illegible] [illegible] гипотез в случае, если переменные являются порядковыми (бывшая глава [illegible]), функциональные схемы "Find the test", ссылки на ресурсы сети Интернет [illegible] ство других материалов.
- *[illegible] /Фонд тестов для преподавателя.* Руководство для преподавателя включает в себя обзор глав, архив вопросов для тестов с множественным выбором, ответы на задачи с четными номерами и пошаговые решения некоторых задач. Кроме того, Руководство для преподавателя включает в себя комплексные упражнения (с ответами), которые можно использовать для тестирования.
- *Путеводитель по курсу.* Путеводитель, написанный профессором Ребеккой Дж[illegible], содержит дополнительные примеры, призванные осветить основные принципы, задачи с подробными ответами для повторения пройденного материала, SPSS-проекты и вопросы для тестов с множественными ответами, дополняющие архив вопросов для тестов, но не дублирующие его.

Обзор ключевых изменений в данном издании

[illegible] наиболее важным изменениям в данном издании относятся:

- [illegible] в их интерпретации, вычислениям на протяжении [illegible] вычислений вручную;
- новый [illegible] под названием "Интерпретация статистики";
- [illegible]
- [illegible] был добавлен MicroCase;
- [illegible]
- набор данных, используемый в материале, был обновлен для его соответствия [illegible]
- менее часто используемые критерии и методы были перенесены на web[illegible]
- многие источники "Чтение статистики" были обновлены.

Материал был тщательно пересмотрен на предмет локализации и [illegible] ти. Как и в предыдущих изданиях, моя цель заключается в том, что [illegible] [illegible]

Благодарности

[illegible]

[illegible] внять совету моих коллег.

Я бы хотел поблагодарить тех преподавателей, которые сделали статистику по- [illegible] Эд Эриксон (Ed Erikson), а также всех моих коллег по Университету Кристофер Ньюпорт за их поддержку и одобрение (особенно профессоров Ф. Самюэла Бауэр[illegible] [illegible] Кернодл (Ruth Kernodle), Тимоти Маршалл (Timothy Marshall), Шелия Мэтьюз [illegible]

[illegible]

1 Введение

Цели изучения

1. [illegible]
2. [illegible]
3. [illegible] переменными, а также приводить примеры таких переменных.
4. [illegible] три шкалы измерений, а также приводить примеры каждой из них.

1.1 Зачем нужно изучать статистику?

[illegible]

Почему так важно учиться управлять данными? [illegible] наиболее важных и передовых работах в области [illegible]

[illegible] В собираемых данных и статистике нет [illegible] ничего не говорит о качестве науч[illegible] ные данные могут оказаться самой надеж[illegible] следовательно, они заслуживают осо[illegible]

[illegible] четко представлять себе [illegible]

[illegible] говорит сама за себя. Исследователь должен уметь эффективно исполь[illegible]

[illegible]

1.2. Роль статистики в научном исследовании

На рис. [illegible] графически показана роль статистики в исследовательском процессе. Диаграмма основана на суждениях Уолтера Уоллеса (Walter Wallace), [illegible] демонстрирует, каким образом база знаний любой научной области растет и развивается. Одна из основных идей диаграммы заключается в том, что научная теория и исследования постоянно влияют друг на друга. Статистика является одним из основных средств коммуникации между исследованиями и теорией. Давайте рассмотрим это колесо более детально.

Рис. 1.1 [illegible]

[illegible]

График является круговым и не имеет ни начала, ни конца, поэтому мы [illegible] будем следовать по стрелкам. Теория представляет собой объяснение связей между явлениями. Людей вечно интересуют проблемы, существующие в обществе (такие как, например, предубеждения, предрассудки, бедность, жестокое обращение с детьми или серийные убийства), и чтобы попытаться понять эти явления, они дают соответствующие объяснения (недостаток образования приводит к предубеждениям). Безусловно, вам знакомо такого рода неформальное теоретизирование относительно общества. Основное различие между нашими неформальными повседневными [illegible] социальных явлений и научной теорией заключается в том, что последняя подвергается тщательному процессу тестирования. В качестве примера давайте возьмем проблему расовых предубеждений и продемонстрируем, как работает исследовательский процесс.

Что является причиной расового предубеждения? Возможное объяснение этого [illegible] сформулировал социальный психолог Гордон Олпорт, и с тех пор она [illegible] реальных проявлений[2]. Теория связывает предубеждение с количеством контактов между членами различных расовых групп. В частности, данная гипотеза утверждает, что контактные ситуации, при которых различные группы имеют одинаковый статус и их поведение основано на принципах сотрудничества, ведут к снижению предубеждения с обеих сторон. Чем выше степень равенства и сотрудничества, тем более вероятно то, что человек будет видеть в другом личность, а не представителя определенной группы. Например, среди членов спортивной команды со смешанным расовым составом, сотрудничающих друг с другом в целях достижения победы, будет наблюдаться тенденция к снижению уровня предубеждений. С другой стороны, когда различные группы борются между собой за рабочие места, жи[illegible] предубеждения.

Гипотеза контакта не является исчерпывающим объяснением предубеждения, но она может послужить примером для иллюстрации роли теории. Мы выявили связь между двумя социальными явлениями: (1) предубеждение, (2) ко[illegible] бой. Те люди, которые имеют меньше контактов между собой, будут [illegible] больший уровень предубеждения друг к другу, и наоборот. Таким образом, мы [illegible] дело с логически последовательной теорией, объясняющей связь между двумя [illegible]

Прежде чем перейти к следующему пункту рис. 1.1, давайте [illegible] на детальном рассмотрении теории. Гипотеза контакта, как [illegible] переменными. Переменная — это свойство, которое может изменять[illegible]

[illegible]

[illegible] Примерами переменных могут служить пол, возрас[illegible] ход или членство в [illegible]ческой партии. В каждой отдельной теории некоторые переменные будут рассматриваться как причины, другие — как следствия или результаты. На научном языке причины называются независимыми переменными, а переменные следствия или результаты называются зависимыми переменными. В нашей теории контакт является независимой переменной (или причиной), а предубежде[illegible] или следствием). Другими слова [illegible]

[illegible] заключаться в том, чтобы дать более конкретные и точные определения [illegible]

[illegible] вательского проекта. Чтобы провести исследование, из теории следует вывести одну или более гипотез. Гипотеза — это утверждение относительно связи между переменными, которое, хотя и логически выводится из теории, однако является гораздо более конкретным и точным.

Например, если бы мы хотели проверить гипотезу контакта, мы должны были бы точно сказать, что мы подразумеваем под предубеждением, и нам бы пришлось очень подробно объяснить, что такое контакт, основанный на равенстве и сотрудничестве. В отношении вопроса влияния контакта на предубеждение было проведено очень [illegible] работки своих определений для этих понятий. По мере того, как наши опре[illegible] следующий уровень исследовательского процесса. Теперь мы рассмотрим этап сбора данных. Нам нужно решить, каким образом можно проверить доводы, как их отбирать, как именно будут измеряться переменные и массу подобных вопросов. В конечном счете эти шаги приведут к фазе наблюдения (низ колеса науки), где мы уже проводим реальное измерение социальной действительности. Прежде чем [illegible]

[illegible] которые бы потребовали от них сотрудничества с членами [illegible] субъектов — в ситуации, в которых происходит конкуренция [illegible] можно было бы сделать путем проведения опроса [illegible] если согласиться или не согласиться со следующими [illegible] не имеет значения, так как люди остаются

[illegible] только независимыми от цвета их кожи". Наша цель заключалась бы в том, чтобы увидеть, действительно ли у людей, оказавшихся в ситуации, включающей контакт с сотрудничеством, произошло снижение уровня предубеждения.

И вот наконец-то мы добрались до статистики! По мере того, как этап наблюдений будет подходить к концу, мы столкнемся с большим количеством численной информации или данных. Если бы наша выборка состояла из 100 человек, мы бы получили 200 заполненных анкет для измерения предубеждения: 100 анкет, заполненных до контактной ситуации, и 100 анкет, заполненных после контактной ситуации. Постарайтесь представить себе работу с 200 заполненными анкетами. Если бы [illegible] задавали каждому респонденту всего пять вопросов с целью измерить его уровень предубеждения, мы бы в итоге получили 1000 отдельных порций информации. Что делать? Нам нужно иметь какой-нибудь способ систематической организации и анализа этой информации, и тогда статистика станет очень ценной. Статистика [illegible] множество идей о том, что делать с данными. Мы перейдем к рассмотрению некоторых возможностей в следующей главе. А пока что, позвольте мне высказать две мысли насчет статистики.

Во-первых, статистика является ключевым элементом. Проще говоря, без статистики невозможно провести исследование количественных характеристик. Без проведения исследований количественных характеристик развитие общественных наук понесло бы серьезный урон. Только путем применения статистических методов обычные данные могут помочь нам формировать и совершенствовать свои теории и лучше понимать социальный мир. Во-вторых, как это ни парадоксально, роль статистики ограничена. Как показывает рисунок 1.[illegible], научное исследование проходит несколько взаимно независимых этапов, и статистика нужна лишь в конце этапа наблюдений. Прежде чем применить какой-либо метод статистического анализа, должны быть успешно завершены предыдущие этапы. Если исследователь задает вопросы, которые [illegible] плохо составлены, или допустил серьезные ошибки в разработке плана или в выборе метода исследования, тогда даже самый совершенный статистический анализ будет бесполезен. Нисколько не умаляя ценности статистики, можно сказать, что она все же не может заменить точного определения понятий, подробного и тщательного планирования или творческого применения теории. Статистика [illegible] плохо составленный или проработанный проект исследования. Она не может извлечь пользу из мусора.

С другой стороны, если применение статистики осуществляется некорректно, это может ограничить полезность проекта, который во всем остальном был выполнен [illegible] вклад в понимание вопроса только путем успешного выполнения всех этапов про[illegible]

По мере того, как объект для статистического анализа [illegible] братиться за эмпирические обобщения. Хотя мы в первую очередь [illegible] ы на оценке своей теории, также можно изучать [illegible] на равенстве и сотрудничестве [illegible] отличается ли это [illegible]

[illegible] и к недостаточно образованным, и к пожилым респонден[illegible] изучения данных мы можем приниматься за рассмотрение [illegible] ий, основанных на наблюдаемых нами эмпирических моделях [illegible] обнаружим, что контакт привел к уменьшению предубеждений [illegible] не оказал влияние на предубеждения пожилых респон[illegible] это означать, что молодые люди менее устоявшиеся в своих взгля[illegible] и отчасти более подвержены изменениям? После разработ[illegible] объяснения мы можем начать исправлять и развивать свою [illegible]

[illegible] мы внесем изменения в теорию, которые бы учитывали эти сведе[illegible] проект исследования, при помощи которого можно было [illegible] ную теорию. Таким образом, колесо науки вновь начинает [illegible] какие-нибудь другие исследователи) вновь пройдем через весь [illegible] ндаемся, более совершенной теорией. Этот второй проект мы [illegible] к дальнейшему исправлению и развитию, которые (по вашему мнению) [illegible] исследовательских проектов, и колесо науки будет продолжать [illegible]

[illegible] вать новые точки зрения. При каждом повороте колеса наше пони[illegible] мого явления будет (выражаем надежду) совершенствоваться [illegible]

[illegible] ученых с пробирками в руках, которые под воздействием ослепительного [illegible] вдохновения находят некоторую фундаментальную истину относительно реальности и кричат "Эврика!" Правда заключается в том, что в науке редко когда можно с абсолютной уверенностью сказать, что данная теория или идея определенно истинна или ложна. Скорее свидетельства в пользу (или против) теорий в конечном счете, со временем накопятся, и окончательные суждения относительно истинности станут результатом многих лет тяжелого труда, исследований и споров.

Давайте вкратце повторно рассмотрим наш воображаемый исследовательский проект. Мы начали с идеи или теории о межгрупповом контакте и расовом предубеждении. Мы представили себе некоторые шаги, которые нам придется с[illegible] чтобы проверить теорию, и вкратце рассмотрели разнообразные этапы исследовательского проекта. Мы вновь оказались на уровне теории, готовые к тому, чтобы начать новый проект, инициированный измененной теорией. Мы видели, как теория может мотивировать исследовательский проект и как наши [illegible] приносят нас к пересмотру теории и, таким образом, мотивировать новый исследо[illegible]

[illegible] между теорией и исследованиями возникает на многих уровнях и во мно[illegible]

[illegible]

[illegible] ния существенно затормозили бы. *В качестве практического задания, касающегося* [illegible]

1.3. Цели книги

В предыдущем разделе я утверждал, что статистика является ключевым элементом процесса, в соответствии с которым выполняются научные исследования, и что по этой причине преподавание статистического анализа является ключевым компо[illegible] в образовании каждого социолога. В этом разделе мы рассмотрим вопросы, [illegible] юшиеся того, насколько необходимо преподавание и какие его цели. Во-первых, этот учебник рассматривает статистические методы как набор инструментов, [illegible] представлять собой очень полезные средства повышения уровня знаний о социальном мире, но они не являются самоцелью. Соответственно, мы не станем заниматься изучением математического подхода к вопросу. Методы будут представлены в виде набора инструментов, которые можно использовать для поиска ответов на важные вопросы. Такой акцент, конечно же, не означает, что мы совсем сможем [illegible] иться без вычислений. Этот материал содержит достаточное количество математического материала, чтобы развить основы понимания того, почему статистика делает то, что она делает. Однако мы будем сфокусированы на том, как применять эти [illegible] ды к общественным наукам.

Во-вторых, все вы вскоре будете выполнять сложную и объемную курсовую работу по своим профилирующим дисциплинам, и вы найдете, что большая часть литературы, используемой в этих курсах, предполагает, по меньшей мере, наличие хот[illegible] основ статистической грамотности. Более того, после получения диплома многие из вас попадут в такое положение (либо в своей работе, либо в аспирантуре), [illegible] некоторое понимание статистики будет очень полезным или даже необходимым. Очень немногие из вас станут статистиками (а этот материал не предназначен для подготовки профессиональных статистиков), но у вас должно быть понимание основ статистики, чтобы вы могли читать и понимать профессиональную литературу по своему предмету. Как студент-социолог вы просто не сможете полностью реализовать свой потенциал без знания статистики.

С учетом этих ограничений данный учебник является введением в статистику в том виде, в котором она используется в общественных науках. Главная цель материала заключается в том, чтобы развить понимание важности статистики и места, занимаемого ею в исследовательском процессе. В результате вы должны будете обрести способность правильно применять статистику и знать, была ли статистика корректно применена другими людьми. Вам следует быть осведомленными о преимуществах и об ограничениях наиболее часто используемых статистических методов и знать, какие методы подходят для каждого конкретного набора данных и конкретной цели. Наконец, вам следует выработать значительные статистические и вычислительные навыки и получить достаточно опыта в интерпретации статистических показателей, чтобы уметь применять некоторые элементарные методы анализа да[illegible]

14. Описательная статистика и статистические выводы

[illegible] стики заключается в том, чтобы [illegible]. Существуют два обш[illegible] [illegible] можно использовать для выполнения этой задачи, и каждый из них представ[illegible] в данном разделе.

Описательная статистика

[illegible] описательной статистики [illegible] описание одной переменной, тогда наша цель будет заключаться в [illegible] таким образом, чтобы можно было быстро понять и оценить релевантную информацию. Многие статистические показатели, которые могли бы быть приемлемы для обобщения, скорей всего вам знакомы. Например, процентные соотношения, графики и диаграммы можно использовать в качестве описательных статистических показателей для одной переменной.

Чтобы вкратце объяснить полезность такого рода статистических показателей, рассмотрим следующую задачу. Предположим, что вы хотите подвести итоги относительно распределения переменной *семейный доход* для выборки из 10000 семей. Как бы вы это сделали? Очевидно, что вас не удовлетворило бы простое перечисление всех показателей дохода по выборке. По всей видимости, вы предпочли бы разработать определенные подытоживающие меры всего распределения дохода — возможно, среднее арифметическое или процентные доли доходов, попадающих в определенные диапазоны (например, низкий доход, средний доход и высокий доход) [illegible] ны не [illegible], он имеет одну задачу — превратить тысячи отдельных элементов [illegible] [illegible] вычисление нескольких чисел, предназначенных для представления множе[illegible] [illegible] сверткой данных, которая представляет собой основную [illegible] [illegible] шена первая часть данной [illegible] [illegible] вать связь между двумя или более переменными [illegible] мерами тесноты связи [illegible] [illegible] они дают нам [illegible] ющих важное теоретическое и практиче[illegible] [illegible] ше [illegible] в зависимости от тесноты связи [illegible] [illegible] Обратите внимание на то, что сами по себе меры тес[illegible] [illegible] что вас интересует связь между временем, потраченным [illegible] изучение статистики, и итоговой оценкой по статистике, и вы собрали соответствующие данные в группе студентов колледжа. Вычислив соответствующую меру [illegible], можно определить силу и направление связи. Предположим, что вы обнаружили сильную положительную связь между этими переменными. Можно сделать вывод, что время на изучение и оценка тесно связаны (степень тесноты связи) и при [illegible] ременной (направление связи). Можно делать предсказания относительно того, каким будет значение одной переменной при определенном значении другой переменной (например, чем больше времени на изучение, тем выше оценка).

Теперь в результате обнаружения такой сильной положительной связи, вы можете поддаться искушению сделать причинно-следственный вывод. То есть, вы можете [illegible] перейти к таким заключениям, как вывод, что большая продолжительность времени [illegible] на изучение ведет к более высоким оценкам. Подобное заключение может [illegible] держать в себе много здравого смысла и оно, определенно, будет подтверждаться статистическим анализом. Однако причинная природа связи никоим образом не доказывается статистическим анализом. Меры связи можно использовать в качестве важного ключа к пониманию причинно-следственной зависимости, но простое существование связи никогда не следует принимать за убедительное доказательство причинно-следственной зависимости.

В действительности, влияние на связь могут оказывать другие переменные. В приведенном выше примере мы бы вряд ли обнаружили безупречную связь между [illegible] на изучение и итоговой оценкой. То есть, возможно, мы столкнемся с некоторыми людьми, которые тратят на учебу много времени, но получают низкие [illegible], равно как и с теми, для которых имеет место противоположная ситуация. Интуитивно мы знаем, что другие переменные, кроме времени на изучение, тоже влияют на оценку (эффективность методов изучения, объем знаний по математике и даже случайность). К счастью, исследователи могут включить в анализ также и эти переменные и измерить их влияние. Часть III материала посвящена двумерным (две переменные), часть IV — многомерным (более чем две переменные) методам [illegible] между переменными.

Статистические выводы

Второй класс статистических методов становится важным, когда нам нужно обоб[illegible] генеральной совокупности [illegible] например [illegible] совокупность — это полностью все объекты исследования, определенную характеристику которых исследователь стремится изучить. Примерами возможной генеральной совокупности могут служить взрослые избиратели-мужчины в Соединенных [illegible] тах, все парламентские демократии, безработные пуэрториканцы, живущие [illegible] те, или футболисты, учащиеся на втором курсе в колледжах Среднего [illegible]

Теоретически размер генеральной совокупности может варьировать [illegible] ятно большого (все человечество) до совсем маленького (все 35- [illegible]

[illegible] генеральные совокупности [illegible] [illegible] [illegible] ...ователей практически никогда не бывает ресурсов или времени [illegible] генеральной совокупности [illegible] возникает необходимость в статистических выводах, которые основаны на [illegible] из выборок тщательным образом отобран- [illegible] [illegible] полученные на этих выборках, могут послужить очень точным представлением о ге- [illegible]

[illegible] телями статистических выводов, особенно нам знакомы такие понятия, как опрос общественного мнения и прогноз [illegible] выборов. Когда мы узнаем, что опрос общественного мнения показал, что 4[illegible] % электората Америки собирается голосовать за определенного кандидата в президенты, по существу это означает, что [illegible] обобщение для всей генеральной совокупности (электорат Америки, насчитывающий около 100 миллионов человек) на основании тщательным образом сформированной выборки (насчитывающей обычно около 1500 респондентов). Мы перейдем к вопросам статистических выводов в части II этой книги *(в качестве при-* [illegible] *различных видов применения ст*[illegible]

1.5. Дискретные и непрерывные переменные

[illegible] человек, и пока мы [illegible] непрерывн[illegible] [illegible] ически. Хорошим примером такой переменной яв-

[illegible] ными значениями. Мы можем указать в отчете, что кто-то пробежал дистанцию в 100 ярдов за 10,7 секунд либо за 10,73245 секунд, но так как время можно делить до бесконечности (вопрос лишь в технологии для проведения точных измерений), мы никогда не сможем указать точное значение прошедшего времени. Так как мы не мо- [illegible] конечно длинные числа (в смысле количества знаков после запятой), и работать с ними, мы должны указать значение непрерывной переменной так же, как если бы оно было дискретным. Различие между двумя видами переменных больше [illegible] осится к измерениям и обработке информации, чем к внешнему виду данных. Это различие между дискретными и непрерывными переменными является [illegible] из основных в статистике, и оно составляет один из критериев, по которым [illegible] между различными статистическими и графическими мето- [illegible] *ческого задания, касающегося различия между дискретными* [illegible] *ыми переменными, см. задачи с 1.4 по 1.8.*

1.6. Шкалы измерений

Вторым основным и общим указанием для выбора статистического показателя является шкала измерений. Каждый статистический метод включает проведение определенных математических операций, таких как сложение значений или ранжирование наблюдений. Прежде чем вы сможете правильно применить метод, вы должны [illegible] те измерить обрабатываемую переменную таким способом, который делает возможным использование требуемых математических операций. Например, многие статистические методы требуют сложения значений. Эти методы можно правильно использовать только в тех случаях, когда переменная измерена таким способом, который допускает сложение. Итак, выбор статистических методов исследователем в значительной степени зависит от того, каким образом была измерена переменная.

Существует три шкалы измерений: номинальная, порядковая и интервальная/относительная (шкалы перечислены в порядке возрастания сложности). Определение шкалы измерений для переменной является одним из первых этапов любого статистического анализа, и мы в дальнейшем рассмотрим этот вопрос более подробно. По всему материалу я буду делиться с вами соображениями по поводу соответствия конкретного статистического метода шкале измерений.

Номинальная шкала измерений

Наиболее простой процедурой измерения является классификация наблюдений по категориям переменной. Все остальные шкалы измерений предполагают, как минимум, возможность выполнения классификации, а для номинальной шкалы классификация по категориям — это *единственная* разрешенная процедура измерения. Категории переменной не являются численными и их можно сравнивать друг с другом только на основе количества наблюдений, относящихся к ним. При этом одни

категории не может быть выше или ниже другой в соответствии с некоторой математической шкалой.

[illegible] переменную, имеющую только две категории, самое [illegible] иметь переменная. Пол можно измерить простым отнесением [illegible]

[illegible]

[illegible] при религиозной принадлежности) сами по себе не формируют математи[illegible] шкалу.

[illegible] ватность измерения. Во-первых, категории переменных, относящихся [illegible] ной шкале, должны быть взаимоисключающими, чтобы не существовало неопределенности относительно классификации любого заданного экземпляра. Во-вторых, категории должны быть исчерпывающими: каждому возможному значению, которое мы можем встретить, должна соответствовать определенная кате[illegible]рия (хотя бы категория прочие).

В-третьих, категории номинальных переменных должны быть срав[illegible] нородными. Это означает, что категории должны включать сопоставимые экземпляры. Другими словами, следует избегать категорий, которые сваливают в одну кучу яблоки и апельсины. Не существует конкретных указаний для оценки того, является [illegible] ответствующее решение с точки зрения конкретной цели исследования и катего[illegible]

Таблица 1.1. Четыре шкалы для измерения религиозной принадлежности

Шкала A не взаимоисключающая	Шкала B не исчерпывающая	Шкала C не однородная	Шкала D однородная
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	Не имеет религиозной [illegible]

[illegible]

[illegible] протестант, которая слишком обширна для многих исследовательских целей. Шкала [illegible] к эвтаназии, абортам или смертной казни), возможно, потребуется сделать [illegible] иудаистов, буддистов, мусульман и т.д.

[illegible]но, иногда для идентификации категорий переменной, измеряемой по [illegible] в случаях, когда происходит подготовка данных к компьютерному анализу. [illegible]мер, различные религии можно закодировать следующим образом: 1 — про[illegible] католик и т.д. Вам следует понять, что эти числа являются лишь метками [illegible] не несут в себе численного значения. Их нельзя складывать, вычитать, умножать или делить. Единственной математической операцией, разрешенной [illegible]менных, является подсчет количества появлений этой переменной в различных категориях.

Порядковая шкала измерений

Вдобавок к классификации элементов по категориям переменные, измеряемые по порядковой шкале измерений, позволяют ранжировать эти категории в зависи[illegible] шкалу, согласно которой их можно упорядочить от большего до ма[illegible]. Таким образом, переменные, которые измеряются по порядковой [illegible] рощные наблюдений. Мы можем не просто сказать, что один экземпляр отличен от другого; мы можем сказать, что одно наблюдение выше или ниже, больше или мень[illegible]

[illegible]

4 [illegible]
3 [illegible]
2 Рабочий класс,
1 Низший класс

[illegible] ...собой на основании категорий. [illegible]

[illegible]

...основное значение характеризуется лишь положением... [illegible] ...значение высокие значения от низких... [illegible] ...ворому другому значению... [illegible] ...в точных единицах. Хотя мы [illegible]

Так как нам не известно... [illegible] ...возможности статистического анализа для нас ограничены [illegible]

[illegible]

...к порядковой шкале. Самой сложной математической операцией, [illegible] ...категорий и наблюдений, хотя, как мы увидим, социологи обычно [illegible] ...вольностей с этим строгим критерием.

Интервальная, относительная шкала измерения

Категории переменных, измеряемых по номинальной шкале, не имеют числен[illegible] ...категории, которые можно упорядочить от высоких до низких величин, но [illegible]

[illegible]

[illegible] ...измерения, которая ...дает интервально-относительные данные, потому что единица [illegible] ...Подобным образом, если мы спросим у человека, сколько у него братьев [illegible]

[illegible] ...переменных является ... [illegible]

[illegible]

[illegible] ...то, что измеряется. Например, переменная *количество братьев и сестер* имеет истинное начало отсчета, потому что возможна ситуация, когда братьев и сестер нет вообще. Подобным образом можно иметь нуль лет образования, [illegible] ...учить нуль баллов по тесту и иметь возраст в нуль [illegible] ...и ненадолго). К другим примерам интервально-относительных переменных относятся количество детей, вероятная продолжительность жизни и количество [illegible] ...в браке.

[illegible] ...подводит итоги этого обсуждения, представляя основные характеристики [illegible] ...измерения. Наиболее важно то, что, так как различные [illegible] ...и требуют применения различных математических операций, вам следует запомнить, что шкала измерения является первым указанием при выборе [illegible] ...показателя. Например, вычисление среднего арифметического [illegible] ...требует сложения и деления, а эти математические операции допустимы [illegible] ...только применительно к переменным, измеряемым по интервальной [illegible]

Таблица 1.2. Основные характеристики трех шкал измерения

Шкала	*Примеры*	*Процедуры измерения*	*Разрешенные математические операции*
Номинальная	Пол, раса, [illegible]	Классификация [illegible]	[illegible] наблюдений в каждой [illegible]
[illegible]	[illegible]	Классификация по [illegible]	Все вышеперечисленное, а также суждения типа "больше" [illegible]
[illegible]	Возраст, [illegible]	Все [illegible] значениями на основе равных единиц	Все вышеперечисленное, а также прочие математические операции (сложение, [illegible] квадратного корня и т.д.)

Чтение статистики 1: Введение

К этому этапу вашего образования вы развили впечатляющее множество навыков, связанных с чтением слов. Хотя иногда вам бывает трудно одолеть сложную концепцию [illegible] ...привести в замешательство неясность значения отдельных слов, все [illegible] ...способны понять практически любой письменный труд, с которым вы можете [illegible]

[illegible] ...набор блоков под названием "Чтение статистики". Они будут встречаться в большинстве глав, и в них будет обсуждаться то, каким образом статистические результаты обычно

[illegible] множество объектов, в которых заинтересован ис- [illegible]

[illegible] между переменными, полученное на [illegible] конкретными, чем теории, все тер- [illegible] определены.

[illegible] часть исследовательского проекта и [illegible]

[illegible] Переменная измерения в единицах измерения, которые [illegible]

[illegible]

[illegible] ох распределения одной переменной; (2) измерением взаимосвя- [illegible]

[illegible] ких методов организации и анализа данных.

[illegible]

[illegible] характеристика переменной, определяемая [illegible] критерий для выбора статистического мето- [illegible] измерены по одной из трех шкал, каждая из кото- [illegible] ение определенных математических операций и стати- [illegible] характеристики этих трех шкал приведены в таблице 1.2.

Мультимедийные ресурсы

[illegible]

Для доступа к широкому спектру студенческих ресурсов посетите [illegible] сопровождающий шестое издание книги *Statistics: A Tool for Social Research*, [illegible] щелкнув на раздел "Студенческие ресурсы" (Student Resources) web-сайта [illegible]

- Обзор основ математики
- Карточки для обучения
- Дополнительные задачи по главам
- Обзор статистики
- Интернет-ссылки
- [illegible]
- [illegible]
- [illegible]
- [illegible]

Задачи

1.1. [illegible] пользуя в качестве основы колесо науки, объясните, каким образом статистика связывает теорию с исследованием.

1.2. Найдите исследовательскую статью в любом журнале, посвященном обществен- [illegible]

а. Какая часть статьи посвящена статистике, как таковой (отдельно от теории [illegible]

б. Основано ли исследование на выборке из некоторой генеральной совокуп- [illegible] рую генеральную совокупность?

в. [illegible] мерения и то, является ли эта переменная дискретной или непрерывной.

г. [illegible] понять. Сохраните статью и прочтите её еще раз по окончании этого курса, чтобы увидеть, насколько улучшится ваше понимание.

1.3. Объясните различие между описательной статистикой и [illegible] выводами. Опишите исследовательск[illegible] каждая из них.

1.4. Ниже приведены нек[illegible]

[illegible]

… четырех названий для ука- [illegible] … качестве характери- [illegible]

[illegible]

в. [illegible] … средний балл? ______

г. [illegible]

д. [illegible] … ый способ борьбы с наркотиками заключается в том, чтобы не [illegible] … все наркотики

[illegible]

1.5. [illegible]

а. Расовая принадлежность. [illegible]

[illegible] Белый ______ Другой

[illegible] … респондентами велось наблюдение в том месте, где лежал предно- [illegible] … терянный кошелек. В кошельке были деньги и полная инфор- [illegible] … ные. Субъекты были отнесены к одной из следующих катего- [illegible]

[illegible]

[illegible] …к, но оставил деньги себе

[illegible]

в. Социальный класс. [illegible] … ов спросили о том, какой была ситуация в их [illegible] … ли их семья

[illegible]

г. **Образование.** Респондентов спросили о том, сколь[illegible] … ч и каждый из их родителей

д. **Взаимопроникновение рас на территории студенческого городка.** В течение месяца за студентами велось наблюдение в кафетерии во время ленча. Подсчитывалось количество студентов, сидящих вместе со студентами других рас

е. **Количество детей.** Респондентов спрашивали: сколько детей у них когда-либо было?

ж. **Расположение студентов в аудитории.** На первом занятии семестра преподаватели отметили, где сидел каждый студент. Измерение расположения студентов проводилось каждые две недели до конца семестра. Каждый студент был отнесен к одной из следующих категорий:

сидел там же, где и при последнем измерении

сидел рядом

сидел в другом месте, но не рядом

отсутствовал

з. **Количество врачей на душу населения.** Было подсчитано количество практикующих врачей для каждого жителя из 50 городов, и исследователи использовали данные о генеральной совокупности для вычисления количества врачей на душу населения

и. **Физическая привлекательность.** Группа, состоящая из [illegible]0 экспертов, оценивала каждую из 50 фотографий выборки со смешанными расами, включая и мужчин, и женщин, на предмет физической привлекательности по шкале от 0 до 20, где 20 представляет наивысшее значение

к. **Количество аварий.** Было записано количество автомобильных аварий на каждом из 20 перекрестков с оживленным движением. Кроме того, каждая авария оценивалась следующим образом:

минимальные повреждения, без травм

средние повреждения, травмы, требующие госпитализации

тяжелые повреждения и травмы.

1.6. Для каждого из первых 20 вопросов Всеобщего общественного опроса (см. приложение Ж) укажите шкалу измерения, а также то, непрерывна или дискретна соответствующая переменная

1.7. Для каждой исследовательской ситуации из тех, что приведены ниже, определите шкалу измерения всех переменных, а также укажите непрерывна или дискретна соответствующая переменная. Кроме того, решите, какие статистические [illegible] … оды описательной статистики для одной перемен- [illegible]

а. [illegible]

[illegible] каждого студента из вашей группы сказать вам, какую [illegible] изучения курса математики, а так[illegible] по статистике. [illegible] не спросите, какая у него оценка за последний тест по статистике [illegible] сравните два набора значений, чтобы увидеть, существует ли между ними [illegible]

[illegible] чтобы вы провели опрос среди избирате[illegible] [illegible] гонорар), чтобы вы провели опрос среди избирате[illegible] [illegible] политике. А именно, они [illegible]

[illegible] «против» [illegible] в какой — против.

[illegible] количестве убийств в штате за двухлетний период до и после изме[illegible] нения.

Местный агент по продаже автомобилей интересуется степенью удовлетво[illegible] рения потребителей. Он хочет всем клиентам за последний год отправить [illegible] форму для опроса и спросить их, удовлетворены ли они, очень удов[illegible] летворены или не удовлетворены своей покупкой.

1.8. Для каждой из представленных ниже исследовательских ситуаций определите зависимые и независимые переменные. Классифицируйте каждую из них с точки зрения шкалы измерения, а также укажите, дискретна или непрерывна данная переменная.

а. [illegible] в связи с чем он опросил 500 студентов [illegible] событий. Каждую из них попросили оценить частоту таких случаев (часто [illegible] никогда). Исследователь также собрал данные о [illegible] дисциплине, чтобы увидеть, существует ли какая-либо [illegible]

б. Инспектор из Отдела обработки твердых отходов городского управления пы[illegible] тается провести оценку двух различных методов сбора мусора. Одна часть [illegible] контейнеров на обочинах дорог. К мерам оце[illegible] количество времени в день, требуемого на обслу[illegible] [illegible] и стоимость тонны собранного мусора.

в. [illegible] расположенных на территории [illegible] [illegible] опрос среди студентов и узнать их [illegible] мнения. В [illegible] у каждого студента [illegible] согласен ли он с утверждением, что [illegible] является [illegible] [illegible]

[illegible] о качестве жизни и степени политической демократии для 50 [illegible] В частности, он использовал уровень детской смертности [illegible] [illegible] качества жизни [illegible]

д. Дорожный инженер интересуется [illegible] уменьшение ограничения скорости на загруженных местных дорогах [illegible] [illegible] количестве аварий и числе [illegible]

е. Студенты планируют провести среди студентов программу [illegible] [illegible] половому сексу и осведомленности о разнообразии [illegible] вопросов [illegible] ния. Чтобы определить эффективность программы, они планируют провес[illegible] ти опрос с целью измерения уровня знаний об этих вопросах в [illegible] выборке студентов до и после программы.

ж. Несколько штатов значительно сократили бюджеты [illegible] [illegible] Увеличит ли это [illegible] [illegible] связаться со множеством агентств, помогающих бездомным в каждом [illegible] [illegible] размер этой генеральной совокупности до и после сокращения [illegible]

Введение в автоматизированные статистические пакеты и Всеобщий общественный опрос

[illegible] количеством наблюдений, чтобы утомительность обычного [illegible] Чтобы [illegible] более ре[illegible] [illegible] общественного опроса 1998 года. Эти [illegible] с некоторыми [illegible] [illegible] http://sociology.wadsworth.com [illegible] Также [illegible] [illegible] 1972 года [illegible] Штатов почти каждый год [illegible] социальных [illegible] литических проблем. Вариант, представленный для этой книги, содерж[illegible] ченное количество переменных и наблюдений, но, тем не менее [illegible] актуальные данные реальной жизни, поэтому у вас есть возможность по[illegible] навыки в области статистики в более реалистичной среде.

Одна из проблем, связанных с реальностью, конечно же, заключается в том, что часто она нестабильна и противоречива. Довольно сложно выполнять домашн[illegible] дание на упрощенных задачах, и вам следует быть очень осмотрительными, с точки зрения времени и затрачиваемых усилий, в вопросах значимости и реализма. [illegible] приводит нас ко второй цели этого раздела: компьютеры и статистические [illegible]. Статистический пакет — это набор компьютерных программ для анализа данных. Преимущество этих пакетов заключается в том, что так как программы уже написаны, вы можете извлечь пользу из мощи компьютера с минимумом компьютерной [illegible] программирования.

[illegible] Package for the Social Sciences (SPSS) и MicroCase. Ваш преподаватель укажет, какой из них вы будете использовать при изучении этого предмета. В этих разделах в конце каж[illegible] [illegible] Прежде чем [illegible] какой-либо анализ данных, прочитайте приложение E.

Часть I

ОПИСАТЕЛЬНАЯ СТАТИСТИКА

2 Основы описательной статистики

Процентные показатели, пропорции и удельные показатели, таблицы, диаграммы и графики

Цели изучения К концу этой главы вы сможете

1. Объяснить роль описательной ст[illegible] части данных
2. Вычислять и интерпретировать проце[illegible] пропорции, отношения и удельные показа[illegible]
3. Строить и анализировать частотные распределения [illegible] ых, относящиеся к каждой из трех [illegible] ал измерен[illegible]
4. Строить и анализировать столбчатые и секторные диаграммы, гистограммы и линейчатые графики

Результаты исследований сами о себе ничего не говоря[illegible] для того, чтобы организовать дан[illegible] Сводка данных подразумевает использование несколь[illegible] ких чисел, таблицы или графика с целью обобщить или дать представление о большом массиве данных.

В этом процессе [illegible]

[illegible]лать. Разнообразные методы подведения итогов дают результаты в различных фор[illegible], основанные на различных математических операциях, и потери информации [illegible] происходят по-разному. Таким образом, при выборе между различными методами подведения итогов следует принять несколько решений: каким образом представлять данные, какую информацию опустить и каким количеством подробностей [illegible]

В этой главе мы рассмотрим несколько распространенных методов представления результатов исследований: процентные показатели, удельные показатели, отношения, пропорции, таблицы, диаграммы и графики.

2.1 Процентные показатели и пропорции

[illegible] Из 269 рассматриваемых судом дел в 167 [illegible] ены к тюремному заключению на пять [illegible] ничего плохого, тот же факт можно было выразить более четко, если бы в нем приводилось процентное содержание: "Около 62% всех дел завершились приговором к тюремному заключению на пять и более лет".

Процентные показатели и пропорции задают систему отсчета для составления отчетов по результатам исследований в том смысле, что они стандартизуют необработанные данные: процентные показатели приводят данные к основанию 100, а пропорции — к основанию 1,00. Математические определения пропорций и процентных показателей формулируются следующим образом:

Формула 2.1

$$\text{Пропорция } (p) = \frac{f}{N}$$

Формула 2.2

$$\text{Процентный показатель } (\%) = \frac{f}{N} \times 100$$

[illegible] во наблюдений в каждой [illegible] ной категории

N — количество [illegible] наблюдений во всех категориях

[illegible] тель наблюдений, относящихся к первой категории (тюремные заключения на пять и более лет), обратите внимание на то, что данная категория содержит 167 наблюдений ($f = 167$), а общее число наблюдений — 269 ($N = 269$). Таким образом,

$$[illegible] \text{ показатель } (\%) = \left(\frac{f}{N}\right) \times 100 = \left(\frac{167}{269}\right) \times 100 = (0{,}6208) \times 100 = 62{,}08\%$$

Используя те же процедуры, мы можем найти также процентный показатель [illegible] дел во второй категории:

$$\text{процентный показатель } (\%) = \left(\frac{f}{N}\right) \times 100 = \left(\frac{72}{269}\right) \times 100 = [illegible]$$

Таблица 2.1. Распределение приговоров по 269 уголовным делам (вымышленные данные)*

[illegible]	[illegible]	Пропорция (p)	Процентное содержание (%)
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0,2677	26,7
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0,0372	[illegible]
[illegible]	[illegible]	1,000	[illegible]

[illegible]

Практический пример 2.1

[illegible] исполнительным директором и инспектором одного из отделов следующий [illegible]

Исполнительный директор — Хорошо, я не хочу выглядеть резкой, но у меня есть всего несколько минут. Расскажите мне, только как можно короче, [illegible] этой проблеме кадрового обеспечения, на которую вы жаловались.

Инспектор — Госпожа директор, у нас просто недостаточно людей для того, [illegible] ющих полное время, только 50 работают в моем отделе. Несмотря [illegible] 16722 дел, которые агентство вело в прошлом году, 6231 вел мой отдел.

Исполнительный директор (сдерживая зевоту) — Очень интересно. Я пере[illegible]

[illegible] Так как он хочет сравнить два набора чисел (численность персонала и объем [illegible]) [illegible] убедительным способом представле[illegible] [illegible] сказал — Только 28,25% персонала [illegible] [illegible] дело с 37,26% [illegible] объема [illegible]

[illegible]

Оба результата можно было бы выразить в виде пропорций. Например, пропорция дел, относящихся к третьей категории, равна 0,0744

$$\text{Пропорция} [illegible] = 0{,}0744$$

Процентные показатели и пропорции проще читать и понимать, чем частоты. Это преимущество становится особенно очевидным при попытке сравнить группы разных размеров. Например, основываясь на информации, представленной в таблице 2.2, скажите, какой колледж имеет большее относительное число экономических специальностей? Так как итоговые списки отличаются существенно, такое сравнение сложно сделать на основании необработанных чисел. Чтобы упростить сравнения, разницу в размерах можно эффективно устранить путем приведения обоих распределений к общему основанию 100 или, другими словами, вычислением процентных показателей для обоих распределений. В таблице 2.3 представлены процентные показатели для тех же данных.

Таблица 2.2. Заявленные специальности в двух высших учебных заведениях (вымышленные данные)

Специальность	Колледж А	Колледж В
Экономические науки	[illegible]	[illegible]
Естественные науки	82	279
Общественные науки	137	188
Гуманитарные науки	93	217
	N = 4[illegible]5	N = 996

Таблица 2.3. Заявленные специальности в двух высших учебных заведениях (вымышленные данные)*

Специальность	Колледж А	Колледж В
Экономические науки	24,82	31,3[illegible]
Естественные науки	19,76	28,0[illegible]
[illegible]	[illegible]	[illegible]
Гуманитарные науки	[illegible]	[illegible]
	N = 100,00% (415)	N = 100,01% (996)

[illegible]

Благодаря процентным показателям, приведенным в таблице 2.3, становится легче [illegible] сходства между двумя колледжами. Колледжу А [illegible]

[illegible]

[illegible]

Некоторые пояснения и указания [illegible]

[illegible]

1. [illegible] предпочтительно сообщать данные [illegible] показатели или пропорции. При небольшом числе наблюдений процентные показатели могут претерпевать значительные изменения при [illegible] тельных изменениях в данных. Например, если вы [illegible] включающих 10 мужчин и 10 женщин (т. е. по 50% каждого [illegible]) [illegible] вите еще одну женщину, распределение процентных показателей изменится и станет равным 52.38% женщин и 47.62% мужчин. Конечно [illegible] мере увеличения числа наблюдений каждое дополнительное наблюдение будет оказывать все меньшее влияние. Если бы мы начинали с 500 мужчин и женщин, а затем добавили еще одну женщину, процентный показатель женщин изменился бы лишь на одну десятую процента (с 50 на 50,1%).
2. Приводя в отчете пропорции и процентные показатели, всегда указывайте и количество наблюдений. Это дает читателю возможность судить об адекватности размера выборки и, напротив, помогает предотвратить [illegible] стике со стороны исследователя. Такие утверждения как "каждый третий опрошенный предпочитает курсы статистики всем остальным курсам" могут оказать впечатление, однако это влияние не потеряет свой блеск, если вы узнаете, что [illegible] опрошено только три человека. *Вам следует очень осторожно относиться к отчетам, в которых не сообщается количество осуществленных [illegible]*
3. Несмотря на тот факт, что при вычислении пропорций и [illegible] [illegible] сительно шкал измерений (см. таблицу 1.2). Процентные показатели и пропорции [illegible] [illegible] деление [illegible] относящихся к отдельной категории перем[illegible] [illegible]

2.2. Отношения и удельные показатели

Отношения и удельные показатели [illegible] характеризующих расп[illegible] [illegible] приведении распределения переменной к осн[illegible] [illegible] вычислении процентных показателей и пропор[illegible] [illegible] делением частоты наблюдений одной кате[illegible]

частоту наблюдений другой категории. Математически отношение можно выразить следующим образом:

Формула 2.3

$$\text{отношение} = \frac{f_1}{f_2}$$

где f_1 = количество наблюдений в первой категории
f_2 = количество наблюдений во второй категории

Для того чтобы проиллюстрировать отношения, предположим, что вас заинтересовали сравнительные размеры различных религиозных конфессий, и вы обнаружили, что в данном городе проживает 1370 протестантских семей и 930 католических семей. Чтобы найти отношение протестантов (f_1) к католикам (f_2), следует разделить 1370 на 930. Результирующее значение отношения равно 1.47, что означает, что на каждую католическую семью приходится 1.47 протестантских семей.

Практический пример 2.2

Посмотрите, каково отношение числа специальностей, принадлежащих к сфере [illegible] наук, из таблицы 2.2 в сравнении с числом социологических специальностей в колледже В? На этот вопрос можно ответить, представив частоты [illegible] будет гораздо проще понять, если они будут выражены [illegible] Отношение числа естественнонаучных специальностей к [illegible]

$$\text{отношение} = \frac{f_1}{f_2} = \frac{[illegible]}{188} = 1.48$$

На каждую социологическую специальность в колледже В приходится 1.48 ес[illegible]

Обратите внимание на то, что отношения могут стать очень экономичным способом выражения относительного преобладания между двумя категориями. Тот факт, что в нашем примере протестанты численно превосходят католиков, очевиден и из необработанных данных. Можно было использовать процентные показатели или пропорции для подведения итогов по общему распределению (т. е. 59.56% семей являются протестантскими, 40.44% — католическими). В отличие от этих методов, от[illegible] [illegible] редством измерения относительной частоты появления одной категории на единицу другой категории. Они точно сообщают нам величину численного превосходства одной категории над другой.

О[illegible] пятой в десятичной дроби. Например, отношение, вычисленное выше, мо[illegible] [illegible] 100 [illegible] [illegible] сравнения в отношениях. [illegible]

[illegible] католиков выразилось бы как [illegible] 1,47 [illegible] выразить как 147:100. *В качестве практического задания, касающегося вычисления и интерпретации отношений, см. задачи 2.1 и 2.2*

Удельные показатели представляют собой [illegible] возможных появлений за некоторую единицу времени. Удельные показатели обычно умножаются на некоторую степень числа 10 для устранения запятой в десятичной дроби. Например, общий удельный показатель смертности в популяции определяется как число смертей в данной популяции за год (действительные появления), разделенное на количество людей в той же популяции (возможные появления) [illegible] число затем умножают на 1000. Формулу удельного показателя смертности можно выразить следующим образом:

Количество смертей за [illegible]

[illegible]

Если бы в городе с населением 7000 человек за год произошло 100 смертей, общий уровень смертности за этот год составил бы

[illegible] 4,29

Другими словами, на каждую тысячу человек за этот год пришлось 4,29 смерти. [illegible] 317 000 человек за год произошло бы 20 краж автомобилей, уровень краж [illegible]

$$\frac{20}{237.000}\times 100.000=(0,0005063)\times 100,000 = 0,63$$

Другими словами, на каждые 100 000 человек за рассматриваемый [illegible] *[illegible] и интерпретации уровней, см. задачи 2.3 и 2.4*

[illegible] мы рассмотрели три метода [illegible]

[illegible] относительной частоты, и поэтому они [illegible] между различными группами и/или [illegible]

Практический пример 2.3

В 1995 году в городе с населением 67000 человек было 2500 новорожденных. В 1985 [illegible] новорожденных. Растет ли рождаемость или падает? Хотя, исходя из приведенных [illegible] рождаемости будет гораздо более очевидной, если мы вычислим рождаемость для каждого года. Как и в случае с общей смертностью, рождаемость обычно умножается на 1000 для устранения запятой в десятичной дроби. Для 1985 года

$$\text{общая рождаемость} = \left(\frac{2\,700}{133.000}\right)\times 1000 = 20,30$$

В 1985 году было 20,30 новорожденных на каждую тысячу жителей города. Для 1995 года

$$\text{общая рождаемость} = \left(\frac{2500}{67\,000}\right)\times 1000 = 14,9[illegible]$$

В 1995 году было 14,97 новорожденных на каждую тысячу жителей города. С помощью этих статистических показателей спад рождаемости выражается более [illegible]

2.3. Частотные распределения: введение

Частотные распределения — это таблицы, предназначенные для обобщения распределения значений переменной путем указания количества наблюдений, содержащихся в каждой категории переменной. Они представляют собой очень полезный и широко используемый способ организации данных и работы с ними. В действительности, построение частотных распределений почти всегда является первым шагом любого статистического анализа.

Чтобы проиллюстрировать пользу частотных распределений и предоставить некоторые данные для примеров, предположим, что консультационный центр университета проводит оценку эффективности своих услуг. При проведении любого реального оценочного исследования было бы собрано множество информации у большой группы студентов, но для простоты примера мы ограничим свое внимание лишь четырьмя переменными и 20 студентами. Данные приведены в таблице 2.4.

Обратите внимание [illegible] огромное число наблюдений), очень трудно выявить какую-либо систему или тенденцию [illegible]

[illegible] представленных таким образом [illegible] в таком формате, который позво[illegible] как выглядит распределение переме[illegible]

Таблица 2.4. Данные исследования консультационного центра

Студент	Пол	[illegible]	[illegible]	[illegible]
[illegible]	Мужской	[illegible]	[illegible]	[illegible]

Ко всем частотным распред[illegible] применимо [illegible] взаимоисключающими. Д[illegible] сформулированы таким образом [illegible] дое наблюдение можно было отнести [illegible] категорию. Этот прост[illegible] принцип применим к построению частотных распределе[illegible]

[illegible] этого правила, существует [illegible] [illegible] шкале). Я укажу вопросы, которые [illegible] [illegible] средствами [illegible] полезные советы. Ка[illegible]

[illegible] При интерпретации разумных решений [illegible]

2.4. Частотные распределения для переменных, измеренных по номинальной и порядковой шкале

Переменные, измеряемые по номинальной шкале. Для номинальных переменных построение частотного распределения обычно является довольно простой задачей. Для каждой категории исследуемой переменной подсчитывается количество наблюдений и промежуточные суммы, а также указывается общее количество наблюдений (N). Таблица 2.5 показывает частотное распределение для пе[illegible] [illegible] центра. В показательных целях в эту т[illegible] [illegible] показать, каким образом наблюдения сортируются [illegible] не будет включен в итоговую форму частотного ра[illegible] Обратите внимание на некоторые другие свойства таблицы. В частности, таблица имеет шапку, категории с ясными названиями (мужской и женский) и [illegible] общим количеством наблюдений внизу столбца частоты. *Эти элемен[illegible] [illegible] включены во все таблицы, независимо от переменной или ее шкалы* [illegible]

Таблица 2.5. Пол респондентов. Исследование консультационного центра

Пол	Отметки	Частота (f)
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

[illegible] и 16 женщин. Этот факт гораздо лучше виден из частотного распределения, чем из [illegible]

[illegible] интересовать различия между [illegible] может понадобиться рассматривать эти две категории как [illegible]

[illegible] положении и не в браке

Таблица 2.6 [illegible] положение респондентов. Исследование консультационного центра

[illegible]	[illegible]	
[illegible]	0	[illegible]

Таблица 2.7 Семейное положение респондентов. Исследование консультационного центра

Положение	Частота	
В браке	[illegible]	[illegible]
Не в браке	[illegible]	[illegible]

Переменные, измеряемые по порядковой шкале

Построение частотных распределений для переменных, измеряемых по порядковой шкале, происходит по той же схеме, что и для номинальных переменных. В таблице 2.8 показано частотное распределение переменной *удовлетворенность* из иссле[illegible] был добавлен столбец процентных показателей по категориям. Такие столбцы повышают читабельность таблицы (особенно при больших выборках), и они являются распространенным дополнением к частотному распределению переменных, принадлежащих к любой из шкал измерений.

Таблица 2.8 Удовлетворенность услугами. Исследование консультационного центра

Удовлетворенность	Частота (f)	Процентный показатель (%)
[illegible]	4	20
[illegible]	4	[illegible]

[illegible] свидетельствует о том, что большинство студентов было либо удовлетворено услугами консультационного центра. Наиболее [illegible] (приблизительно для половины выборки) был ответ [illegible] нужно выделить [illegible] потеря [illegible]

Таблица 2.9 Удовлетворенность услугами. Исследование консультационного центра

Удовлетворенность	Частота (f)	Процентный показатель
[illegible]	[illegible]	[illegible]
Неудовлетворен	7	35
	$N = 20$	100%

2.5 Частотные распределения для переменных, измеряемых по интервальной/относительной шкале

Основные соображения

В целом, построение частотных распределений для переменных, измеряемых по интервальной/относительной шкале, является более сложным, чем для номинальных и порядковых переменных. Интервальные/относительные переменные обычно [illegible] значений (т. е. имеет место широкий диапазон значений между наименьшим и наибольшим значениями). Большое число значений требует некоторого упрощения и группирования категорий с целью создания достаточно компактных частотных распределений. Для того чтобы построить частотные распределения для интервальных/относительных переменных, вы должны решить, сколько категорий нужно использовать и насколько широкими должны быть эти [illegible]

Например, предположим, что вы захотели показать распределение переменной *возраст* для выборки, отобранной из жителей города. В отличие от данных, собранных по колледжу, представленных в таблице 2.4, выборка из жителей города будет соответствовать очень широкий диапазон возрастов. Если вы просто укажете, сколько раз встречается каждый возраст (по годам), у вас может быстро получиться 70, 80 и даже больше категорий. Частотные распределения таких больших размеров не дадут законченной картины. Чтобы повысить ясность и упростить понимание значения (годы возраста) нужно сгруппировать в категории больших размеров. Насколько большими должны быть эти категории? Сколько категорий следует включить в таблицу? Хотя не существует непреложных правил для принятия таких решений, вы всегда должны выбирать золотую середину между высокой степенью детализации (большим количеством узких категорий) и высокой компактностью (меньшим количеством широких категорий).

Практический пример 2.4

В следующем списке приведен возраст 50 заключенных, [illegible] ме, где они освобождаются из-под стражи на время работы. [illegible] молодой или старой? Частотное распределение даст нам [illegible]

[illegible]

30	45	67	41	
37	47	65	42	
22	8	27	[illegible]	
[illegible]	23	52	35	42

Для представления этих данных мы будем использовать [illegible] [illegible] что самому молодому заключенному 18 лет, а самому пожилому 75 лет. Таким образом, диапазон [illegible] интервала будет равен 57/10, т.е. 5,7, что можно [illegible] либо до значения 6. Мы будем использовать шестилетний интервал, начинающийся с 18. Пределы начального интервала будут 18–[illegible]. Теперь [illegible] пределы всех остальных интервалов, подсчитать количество заключенных в каждом интервале и вывести эти значения в виде частотного распределения. Можно добавить столбцы процентных показателей, кумулятивных процентов и/или кумулятивной частоты. Полное распределение с добавлением столбца для [illegible]

Возраст	Частота	Доля в процентах
[illegible]	[illegible]	[illegible]
48–53	4	8
54–59	2	4
60–65	[illegible]	[illegible]
66–71	[illegible]	2

Мы видим, что заключенные [illegible]

Построение частотного распределения

Чтобы [illegible] в эти категории, часто [illegible] частотного распределения [illegible], кто моложе, до тех, кто старше [illegible]

[illegible] в соответствие то, сколько раз каждый возраст встречается в выборке [illegible] 2.10 представлена информация [illegible] [illegible] в интервалах группирования 18 и 19.

Таблица 2.10. Возраст респондентов. Исследование консультационного центра (ширина интервала = один год возраста)

Интервалы группирования	Частота (f)
[illegible]	[illegible]
	N = [illegible]

[illegible]

Таблица 2.11. Возраст респондентов. Исследование консультационного центра (ширина интервала = два года)

Интервалы группирования	Частота (f)	Процентный показатель
20–21	5	25
[illegible]	[illegible]	[illegible]
	N = 20	

Обратите внимание на то, что интервалы группирования в таблице 2.11 были [illegible] тельности эти разрывы не представляют проблемы [illegible] мерения (в годах, а не в десятых или сотых частях года). [illegible] может иметь значение, которое бы попало между этим интервалом [illegible]

[illegible] 2.11, составляет непрерывающую и взаимоисключающую шкалу [illegible] из 20 респондентов выборки можно отнести к одной и только одной возрастной категории.

Тем не менее давайте рассмотрим, какие трудности могли возникнуть, если бы [illegible] точностью [illegible] в десятых частях года, и какие [illegible] Этой неопределенности можно было бы [illegible] отнести человека с возрастом 19,4 лет? Этой неопределенности можно было бы избежать указанием пределов интервала группирования с той же т[illegible] рой были записаны данные. Таким образом, если возраст измерялся в деся[illegible] тых года, пределы интервалов группирования в таблице 2.11 можно было бы указать в десятых частях года. Например

17,5–19,4

[illegible]

21,5–23,4

23,5–25,4

25,5–27,4

Чтобы поддерживать взаимное исключение между категориями, *не допускаются взаимные перекрытия интервалов группирования*. Если вы указываете пределы интер[illegible] группирования на том же уровне точности, что и данные, и не допускаете раз[illegible] между интервалами, вы всегда будете получать такое частотное распределение, [illegible] котором каждое наблюдение можно было бы отнести к одной и только одной [illegible]

Средние точки

[illegible] вам понадобится работать со средними точками интервалов группирования [illegible] при построении и интерпретации некоторых графиков. Средние [illegible] которые расположены точно посередине между верхним и ниж[illegible] [illegible] наборов интервалов группирования. *В качестве приме[illegible] нахождения средних точек, см. задачи 2.86 и 2.[illegible]*

Кумулятивная частота и кумулятивный процентный показатель

[illegible] кумулятивной частоты [illegible] возможность с первого взгляда понять [illegible] значение меньше определенного значения или [illegible] деления [illegible] значениям) и рас[illegible] интервала будет таким же

как и количество наблюдений в данном интервале. Для следующего по возрастанию интервала кумулятивная частота будет равняться сумме количества всех наблюдений в этом интервале и количества наблюдений в первом интервале. Для третьего интервала кумулятивная частота будет равняться сумме количества всех наблюдений в этом интервале и количества наблюдений в первых двух интервалах. Продолжайте складывать (или накапливать) наблюдения до тех пор, пока вы не достигнете последнего интервала группирования, для которого кумулятивная частота будет равна сумме количества всех наблюдений в этом интервале и количества наблюдений во всех остальных интервалах. Для последнего интервала кумулятивная частота равняется общему числу наблюдений. В таблице 2.13 приведена таблица 2.[illegible] с добавленным столбцом кумулятивной частоты. Стрелки показывают направление сложения

Таблица 2.12. Средние точки

Ширина интервалов группирования, равная трем единицам	
Категории	Средние точки
0–2	1
1–5	[illegible]
[illegible]	[illegible]

Ширина интервалов группирования, равная шести единицам	
Категории	Средние точки
00–105	102,5
06–1[illegible]	[illegible]8,5
12–117	[illegible]4,5
[illegible]8–123	120,5

Таблица 2.13. Возраст респондентов. Исследование консультационного центра (ширина интервала = два года)

Интервалы группирования	Частота	Кумулятивная частота
18–19	11	11
20–21	5	16
22–23	2	18
24–25	1	19
[illegible]	[illegible]	[illegible]

Столбец кумулятивного процентного показателя очень похож на столбец кумулятивной частоты. Начнем с добавления к таблице частотного распределения столбца процентного показателя, как в таблице 2.1[illegible]. В этом столбце представлен про[illegible] [illegible] найти кумулятивные процентные [illegible] жения, которая была описана [illegible] ный процентный [illegible]

казаны данные о возрасте с добавлением столбц[illegible]

Таблица 2.14. Возраст респондентов. Исследование консультационного центра

Интервалы группирования	Частота *f*	Накопленная частота	Процентный показатель (%)	Кумулятивный процентный показатель
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	6	[illegible]	[illegible]
[illegible]	[illegible]	9	[illegible]	[illegible]
[illegible]	[illegible]	20	[illegible]	[illegible]

Столбцы, в которых представлены накопленные значения, очень полезны в ситуациях, когда исследователь хочет указать на то, как наблюдения распределены по диапазону значений. Например, таблицы 2.13 и 2.14 довольно четко показывают, что большинство студентов в исследовании консультационного центра относились к возрасту младше 21 года. Если исследователь хочет продемонстрировать это свойство возрастного распределения своей аудитории, то столбцы накопленных значений окажутся довольно полезны. В наиболее реалистичных исследовательских ситуациях рассматривается более 20 наблюдений и/или намного больше категорий, чем в наших таблицах. Так как в таких случаях столбец кумулятивного процентного показателя читабельнее и проще в интерпретации, обычно его применение [illegible] чем применение столбца кумулятивной частоты.

Процедуры построения распределения частот для интервальных/относительных переменных

[illegible]

[illegible] мера.

4. Определите пределы интервалов группирования с таким же уровнем точности, который использовался при измерении данных. Не допускайте наложения интервалов. Таким образом вы получите такие интервалы, что каждое наблюдение можно будет поместить в одну и только одну категорию.
5. Подсчитайте количество наблюдений в каждом интервале [illegible] эти промежуточные суммы в столбце "Частота". Внизу столбца укажите общее количество наблюдений (N). Таблица также может включать столбцы процентных показателей, кумулятивных частот и кумулятивных процентных показателей.
6. Тщательно изучите распределение частот. Не было ли потеряно слишком много подробностей? Если это так, перестройте таблицу таким образом, чтобы в ней было больше интервалов группирования (или меньший размер интервала). Не слишком ли много подробностей содержит ваша таблица? Если это так, перестройте таблицу таким образом, чтобы в ней было меньше интервалов группирования (или используйте более широкие интервалы). Помните, что результирующее частотное распределение является следствием достаточно большого количества решений, которые вы принимаете произвольным образом. Если внешний вид таблицы выглядит не очень оптимальным с точки зрения целей исследования, переделывайте таблицу до тех пор, пока вы не будете удовлетворены достигнутым балансом между ее степенью детализации и лаконичностью.
7. Не забудьте дать таблице четкий и лаконичный заголовок и, если ваш отчет содержит больше одной таблицы, пронумеровать ее. Все категории и столбцы также должны иметь четкие заголовки. *В качестве практического задания, касающегося построения и интерпретации частотных распределений для интервальных/относительных переменных, см. задачи 2.5—2.9.*

2.6. Построение частотных распределений для интервальных/относительных переменных: обзор

В предыдущем разделе мы рассмотрели значительное количество материала; поэтому давайте сделаем паузу и вновь повторим изученные принципы, рассмотрев [illegible] количество посещений [illegible] за прошедший год для 90 жителей дома престарелых.

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	24	[illegible]	[illegible]	[illegible]	[illegible]
6	50	40	28	36	12	47	1	[illegible]	[illegible]
9	26	46	52	27	10	3	0	24	50
24	19	22	26	26	50	[illegible]	[illegible]	[illegible]	[illegible]
23	51	18	22	17	24	17	8	28	[illegible]
20	50	25	50	18	52	[illegible]	[illegible]	[illegible]	[illegible]
32	0	24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]	[illegible]	[illegible]		[illegible]	49	42	6	28	2
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	16	50

[illegible] от которой [illegible] о распределении частот [illegible] организовать эти данные таким образом, что [illegible]

[illegible] интервалы [illegible] группировки следует [illegible] интервалов. Изучив данные, мы можем увидеть, что наименьшее значение равно 0, а наибольшее — 52. Диапазон этих значений равен 52 – 0, т.е. 52. Чтобы найти приблизительный размер интервала, разделим диапазон (52) на количество интервалов (10). Так как 52/10 = 5,2, мы можем установить размер интервала равным 5.

Таблица 2.15. Количество визитов в год, 90 жителей дома престарелых

Интервалы группирования	Частота (f)	Кумулятивные частоты	Процентный показатель (%)	Кумулятивный процентный показатель (%)
[illegible]	[illegible]	0	11,11	11,11
[illegible]	[illegible]	15	5,56	16,67
[illegible]	[illegible]	23	8,89	25,56
[illegible]	12	35	15,33	38,89
[illegible]	[illegible]	53	20,00	58,89
[illegible]	[illegible]	65	13,33	72,22
[illegible]	[illegible]	68	3,33	75,55
[illegible]	[illegible]	70	2,22	77,77
[illegible]	[illegible]	72	2,22	79,99
[illegible]	[illegible]	78	6,67	86,66
[illegible]	[illegible]	90	13,33	99,99
[illegible]			99,99%*	

[illegible] плюс-минус 0,10%, это может указывать на [illegible] весь столбец.

[illegible] равно 0, [illegible] для того, чтобы указать интер[illegible] значений, попадающих в каждый [illegible] частоты. Эти шаги были предприняты в [illegible] кумулятив[illegible] [illegible] распределение [illegible] можно попытаться выявить в [illegible] группы значений.

Десятерым жителям было [illegible] 18 [illegible] 42 [illegible] 24 [illegible] 50 – 54 содержащий 2 [illegible] 58,8% [illegible]

Чтение статистики: частотные распределения и диаграммы

По мере повышения уровня использования профессиональной социологической литературы у вас возникнет нужда в том, чтобы интерпретировать и понимать таблицы, диаграммы и графики. Эти графические средства будут объяснены в тексте данной главы, но вам следует развивать навыки самостоятельного чтения и делать свои собственные выводы. Ниже представлены некоторые мысли, которые следует иметь в виду при чтении таблиц и диаграмм.

Во-первых, существует множество различных форматов представления результатов, и те таблицы и графики, которые вы найдете в исследовательской литературе, необязательно будут соответствовать правилам, которые использовались для данного материала. Во-вторых, из-за ограничения места, таблицы и графики будут представлены с минимумом подробностей. Например, исследователь может представить частотное распределение, которое будет содержать только столбец процентного показателя.

Начинайте свой анализ с прочтения заголовка, всех подписей (т.е. заголовков строк и столбцов) и всех сносок таблицы. В результате вы получите точное представление о том, какая информация содержится на рисунке или в таблице. Изучите содержимое таблицы или графика, помня выводы автора. Подумайте, согласны ли вы с выводами и результатами анализа автора (почти всегда вы будете согласны, но никогда не будет лишним его перепроверить и потренировать свои критические способности).

В конечном итоге, помните, что в своем большинстве исследовательские проекты занимаются анализом взаимосвязей между многими переменными. Таблицы, графики и диаграммы, описанные в этой главе, показывают по одной переменной за раз. Исследовательские отчеты вряд ли будут содержать что-то подобное (возможно, только лишь в качестве вводной информации). Даже если это не указано в отчете, вы можете быть уверены в том, что исследование начиналось с изучения частотных распределений или графиков всех переменных. Одномерные таблицы и графики показывают много информации о переменных в компактном, простом для понимания, формате и почти всегда используются в качестве средств описания.

Статистика в профессиональной литературе

В этом разделе представлены две краткие выдержки из социологических исследований, так что вы сможете увидеть, как на практике интерпретируется и указывается в отчетах статистическая информация. Первая выдержка представляет частотное распределение, а вторая — линейчатую диаграмму.

Выдержка 1. Власть и пол

[illegible] реакий, столь же ограниченными в [illegible] ют меньшей степенью [illegible]

[illegible] ерба (Sidney Verba) опросили 380 семейных пар [illegible]

[illegible] Затем данная информация, относящаяся к вопросу [illegible] таблице на стр. 67–68, воспроизведенных точно по оригиналу [illegible] некоторые полученные сведения.

[illegible] ие на то, что эта таблица представляет 10 различных частотных рас- [illegible] процентные показатели и что таблицы расположены горизонталь- [illegible] как в данной книге. Как всегда, исследователи указывают суммы [illegible] и действительное количество наблюдений [illegible] щии, вот что сказали о таблице авторы.

> [illegible] показывает распределение... ответов для жен и мужей. В целом указа[illegible] разделение обязанностей, где жены делают больше работы, связанной с покупкой товаров, заботой о детях и, в меньшей степени, занимаются оп[illegible] а мужья отвечают, в основном, за автомобиль... Доля семейных пар, [illegible] дали противоречивые ответы (каждый заявлял, что он выполняет весь [illegible] конкретной семейной обязанности или большую его часть), относительно [illegible] шее число противоречивых ответов было получено в связи с тради[illegible] ми обязанностями: забота о детях и уборка дома. Больши[illegible] лишь немного мужей, заявляют, что они выполняют весь объем работ, от[illegible] данной обязанности или большую его часть.

[illegible] Nancy, Kay Schlozman, and Sidney Verba, 1997. "The [illegible]

Таблица [illegible] Распределение домашних обязанностей

	Весь объем	Большую часть	Некоторую часть	Малую часть	Ничего		N	Противоречивые ответы
[illegible]								

	Весь объем	Большую часть	Некоторую часть	Малую часть	Ничего		N	Противоречивые ответы
Мужья говорят, что они делают	8	13	38	25	15	100%	349	
Оплата счетов								
Жены говорят, что они делают	46	7	13	12	11	101%	374	9%
Мужья говорят, что они делают	21	20	14	20	26	100%	349	
Ремонт автомобиля								
Жены говорят, что они делают	7	13	29	27	24	100%	374	17%
Мужья говорят, что они делают	69	24	5	1	1	100%	349	17%
Забота о детях								
Жены говорят, что они делают	11	63	22	1	4	101%	330	5%
Мужья говорят, что они делают	—	7	75	10	9	101%	224	

[illegible]

Хотите [illegible] к какому выводу пришли исследователи? Со[illegible] ведена на странице 67, а статью, скорее всего, можно найти в библиотеке.

Выдержка 2. Расизм в детской литературе

Как [illegible] были представлены в детских книгах? Отражаются [illegible] ражениях предубеждения, конфликты и отторжения, имевшие место в обществе? Чтобы это выяснить, исследователи Берниc Пескосолидо (Bernice Pescosolido), Элизабет Гр[illegible] (Elizabeth G[illegible]) и Мелисса Милки (Melissa Milkie) [illegible] тысячи детских книг, опубликованных в Соединенных Штатах между 1930 и 1990 годами. На линейчатой диаграмме, расположенной ниже, показаны некоторые из их результатов.

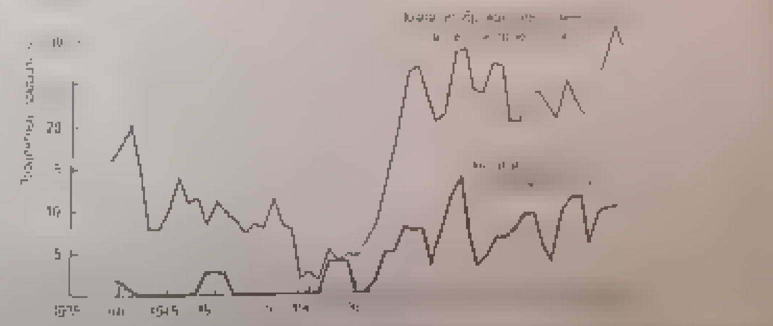


Рис. 1. Процентные показатели [illegible]

[illegible] … в детских книгах систематически [illegible] … появление чернокожих в 3 [illegible] … чернокожие представлены в [illegible] … отмечается практически полное отсут[illegible] … В третьей фазе с середины 1960 до начала 1970 годов [illegible] … с 1975 по 1993 годы, процентный показатель [illegible] … книг в которых изображен [illegible] … колеблется между 20 и 30 процентами.

[illegible] … межрасовых отношений в [illegible] … периодов. Хотите [illegible] к каким выводам они пришли? Ссылка приведена ниже.

[illegible] *Conflict. The Portrayal of Blacks in U.S. Children's Picture Books Through the Mid- and Late-* [illegible]

2.7. Диаграммы и графики

[illegible]

[illegible] … они сравнительно просто в использовании, и [illegible] вам доступны, следует с ними познакомиться. Те усилия, кото[illegible] затратить для изучения подобных программ, будут вознаграждены [illegible]. Раздел в конце этой главы, посвященный компью[illegible] демонстрирует, как создавать столбчатые и линейчатые ди[illegible]

Секторные диаграммы

[illegible] секторные диаграммы [illegible] … каждую категорию перем[illegible] … Затем [illegible]

[illegible] в которой изображено распределе[illegible] … мы выделим 180° (или 50%) для первой категории, 126° (35%) — для второй и 54° (15%) — для последней категории. Визуально на секторной диаграмме усилено от[illegible] … [illegible]ведённых студентов в исследовании консультационного центра.

Рис. 2.1
Пример секторной диаграммы [illegible] *респондентов (N = 20)*

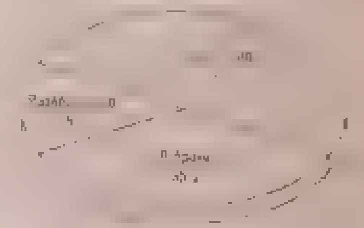


Таблица 2.16. Семейное положение респондентов. Исследование консультационного центра

Положение	Частота (f)	Процентный показатель (%)
Одинок	[illegible]	[illegible]
В браке	[illegible]	[illegible]
Разведен	[illegible]	[illegible]
N = 20		[illegible]

Столбчатые диаграммы

Как и секторные диаграммы, **столбчатые диаграммы** являются сравнительно простыми в понимании. Условно принято, что категории переменной должны откладываться вдоль горизонтальной оси (или абсциссы), а частоты или процентные показатели, если нужно, откладываются вдоль вертикальной оси (или ординаты). Для каждой категории переменной постройте (или нарисуйте) прямоугольник постоян[illegible] … тегории. Столбчатая диаграмма, изображенная на рис. 2.2, воспроизводит данные о семейном положении, отраженные на рис. 2.1 и в таблице 2.16.

[illegible]

...4" и "от 50 до 54"

Рис. 2.2 ... столбчатой диаграммы. Семейное положение респондентов (N = 20) ...

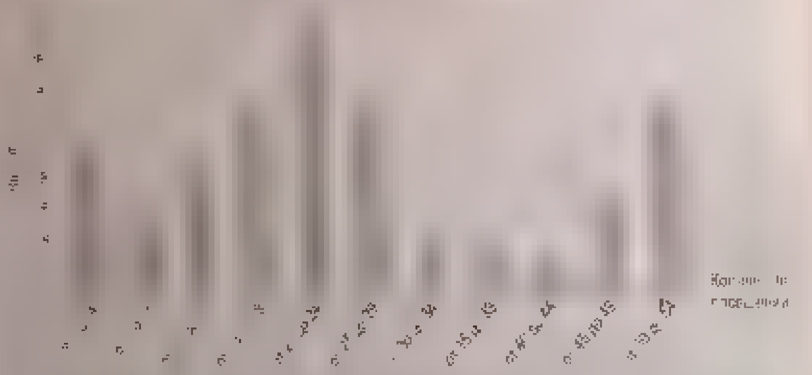


Рис. 2.3 ...

... ает данные эффективным и доступным обра... ...ает, что для мужчин этот уровень выше, чем для ... полов имели наивысшее значение в 1975 году, и ... в этот уровень рос быстрее, чем для жен-...

Гистограммы

Гистограмма ...

... столбцы расположены рядом друг с другом. Для того чтобы построить гистограмму на основе частотного распределения, нужно выполнить следующие шаги.

1. Расположить интервалы группирования или отдельные значения на горизонтальной оси (абсциссе), указывая пределы границ интервалов.
2. Расположить частоты на вертикальной оси (ординате).
3. Для каждой категории частотного распределения построить столбец, высота которого соответствовала бы количеству наблюдений, относящихся к данной категории, а ширина — границам интервалов группирования.
4. Сделайте подпись к каждой оси графика.
5. Дайте графику ...

В качестве примера на рис. 2.5 изображена гистограмма для выбора распределения возрастов выборки респондентов в национальном опросе общественного мнения. График показывает, что в выборке наблюдается концентрация людей в возрасте 30–50 лет и что количество респондентов уменьшается с увеличением возраста. Обратите также внимание на то, что в выборке отсутствуют люди, возраст которых меньше 18 лет, так как этот возраст является точкой отсева респондентов для опросов общественного мнения.

Рис. 2.4 ...

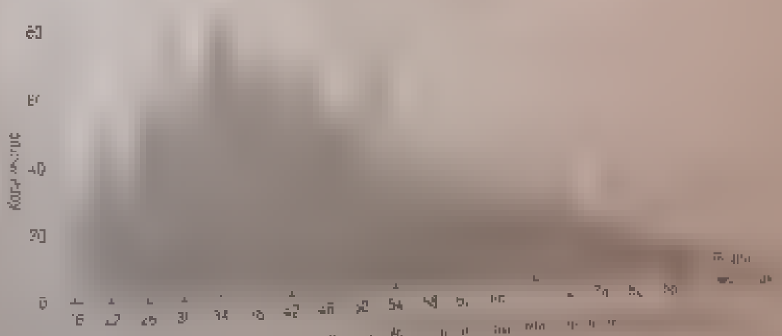


Рис. 2.5 ...

Линейные диаграммы

…частотного полигона похоже на …, того, чтобы для представления … каждого интервала наносится риска. Затем … 2.6 начерчена линейчатая диаграмма … рис. 2.3

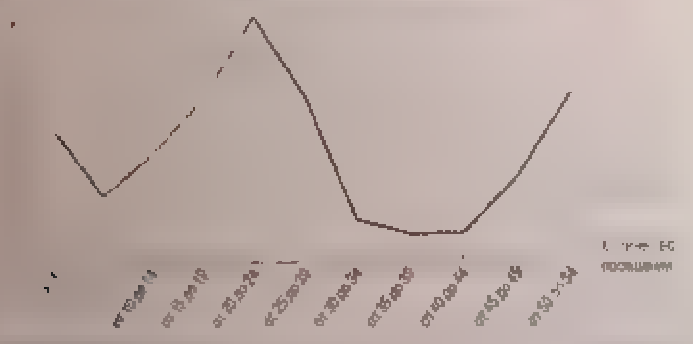
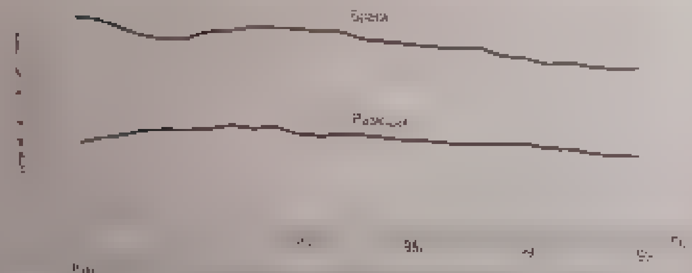


Рис. … Линейчатая диаграмма. Количество посещений в год. Жители дома …

… также можно использовать для изображения временных … показаны удельные показатели браков и удельные показате… 1000 человек населения Соединенных Штатов, начиная с 1972 го… на то, что с середины 1980-х годов значения обеих показателей …



2.8. Интерпретация статистик. Использование процентных показателей, частотных распределений, диаграмм и графиков для анализа изменения степени надзора за рабочим местом

В этой главе был охвачен большой объем статистического материала, и в качестве итоговых рассуждений будет полезно сконцентрироваться на содержательности и интерпретации результатов. Что можно сказать после того, как были вычислены процентные показатели, построены частотные распределения, графики или диаграммы? Помните, что статистика является инструментом, предназначенным для того, чтобы помочь вам анализировать информацию и отвечать на вопросы. Статистические показатели никогда не говорят сами за себя, и их всегда нужно понимать в контексте конкретного вопроса исследования или проверки гипотезы. В данном разделе будет представлен пример интерпретации, где будут поставлены определенные вопросы социологического исследования и получены на них ответы. Интерпретация (или словесное описание) будет явным образом связана со статистическими методами и числовыми данными, так что вы сможете увидеть, как и почему делают …

Ваша новая работа и надзор за рабочим местом

Поздравляем! Вы только что получили работу в крупной американской корпорации, и вот вы сидите в своем кабинете в обеденный перерыв. Может быть, вам стоит зайти на web-сайт D&D и быстро сыграть в свою любимую ролевую игру? А может отослать письма своим старым друзьям по колледжу через систему электронной почты компании или записаться к парикмахеру по телефону, который стоит на вашем столе? Прежде чем принять решение, рассмотрите серию отчетов Американской ассоциации менеджмента (American Management Association, AMA). Эти отчеты сообщают, что растет вероятность того, что вы можете столкнуться с надзором за рабочим местом и что за вашей электронной почтой, телефоном, а в последнее время и за пользованием сетью Интернет может следить ваш работодатель.

Общие тенденции

С 1997 года AMA провела исследования, касающиеся надзора за рабочим местом и слежения за сотрудниками[1]. Исследования затронули компании, являющиеся … [illegible]

… озаглавлена и имеет все подписи. Каждый столбец … [illegible]

[illegible] процентные показатели (не частоты), [illegible]

Из таблицы видно постоянное [illegible] с 1997 [illegible] В этот период [illegible] более чем три четверти [illegible] в том, что [illegible]

Таблица 2.17. Слежение и надзор за рабочим местом, 1997–2000

Занимается ли ваша компания слежением?	Год			
	1997	1998	1999	2000
[illegible]	63,4	67,1	67,3	78,[illegible]
[illegible]	36,6	32,9	32,7	22,6
[illegible]	100,0	100,0	100,0	100,0
	N = 906	1085	1054	2133

Слежение и надзор в 2000 году

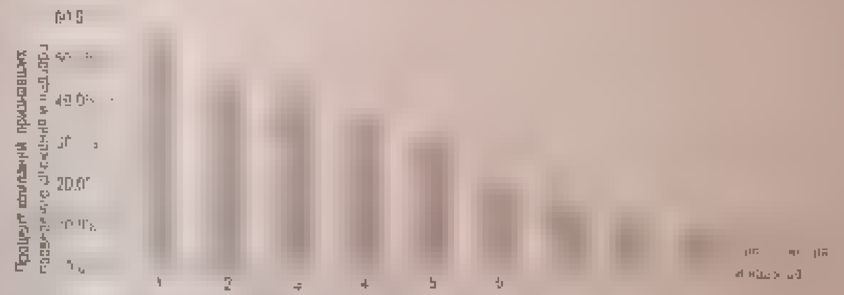
Хотя компьютер становится все более важным компонентом рабочего оборудования для все большего количества людей, ваш работодатель может считать необходимым следить за его использованием. В таблице 2.18 представлен процентный показатель компаний, признающих, что они практиковали ту или иную форму [illegible]ния или надзора.

Таблица 2.18. Слежение и надзор в 2000 году

Тип слежения и надзора	Процентное содержание	
	Да	Нет
[illegible]	[illegible]	[illegible]
[illegible]	44,[illegible]	[illegible]
[illegible] сообщений электронной почты	3[illegible]	6[illegible]
Видеонаблюдение в целях безопасности	35	6[illegible]
Хранение и просмотр компьютерных файлов	30,8	[illegible]
[illegible] компьютера (время начала работы, количество введенных строк и т.д.)	[illegible]	[illegible]
[illegible] производительности труда сотрудника	9,4	[illegible]
Запись и прослушивание телефонных разговоров	[illegible]	[illegible]
[illegible]	[illegible]	88,5
[illegible]	[illegible]	9[illegible]

[illegible] и не формируют шкалу), а так [illegible] диаграмма будет предпоч[illegible] что слежение за использова[illegible]

[illegible]ем сети Интернет было наиболее распространенной [illegible] электронной почты (38,1%). Этот график ясно показывает, что неразумно пользоваться Интернет в рабочее время или использовать телефон или электронную почту [illegible]



Ключ для типов слежения и надзора

1. [illegible]
2. Использование телефона [illegible]
3. Хранение и просмотр сообщений электронной почты
4. Видеонаблюдение в целях безопасности
5. [illegible]
6. Использование компьютера (время начала работы, количество введенных строк, и т.д.)
7. [illegible]
8. [illegible]
9. [illegible]

Рис. 2.8. *Слежение и надзор в 2000 году*

Слежение и надзор в различные годы

Какие изменения происходили в применении конкретных методов слежения с 1997 по 2000 год? Таблица 2.19 показывает, что в 1997 году компании в значительной мере (около 34%) использовали слежение за использованием телефона и видеонаблюдение в целях безопасности. Многие формы слежения и надзора за рабочим местом практиковались лишь в 10–15% компаний, участвовавших в исследовании. [illegible] голосовой почты.

К [illegible] году процентный показатель компаний, отслеживающих использование [illegible]

[illegible]

Таблица 2.4. Слежение и надзор с 1997 по 2000 год

[illegible] по слежению и надзору	Процентный показатель применявших осуществление слежения и надзора			
	1997	1998	1999	2000
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	5,3	[illegible]	6,6	[illegible]
[illegible]	[illegible]	19,6	[illegible]	30,8
[illegible]	[illegible]	2[illegible],2	27,[illegible]	38
[illegible]	[illegible]	5,6	6	6
[illegible] (затраченное время, номера телефонов)	34,4	40,2	38,6	44,0
[illegible] пользования компьютера (время начала работы, [illegible])	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

И снова эти тенденции можно было бы более четко выразить в форме графика. [illegible] как нас интересуют изменения процентных показателей во времени, подхо[illegible] [illegible] линейный график. Рис. 2.9 показывает относительную [illegible] одних форм надзора (например, видеонаблюдения в целях безопасно[illegible]) [illegible] значительный рост других (например, слежение за пользованием телефоном [illegible]) [illegible] самой очевидной тенденцией, отраженной на этом графике, является непре[illegible] рост доли компаний (в процентах), отслеживающих сообщения электрон[illegible] почты сотрудников. В 1997 году только около 15% компаний применяли эту [illegible] слежения и надзора. К 2000 году этот процент более чем удвоился и достиг [illegible]

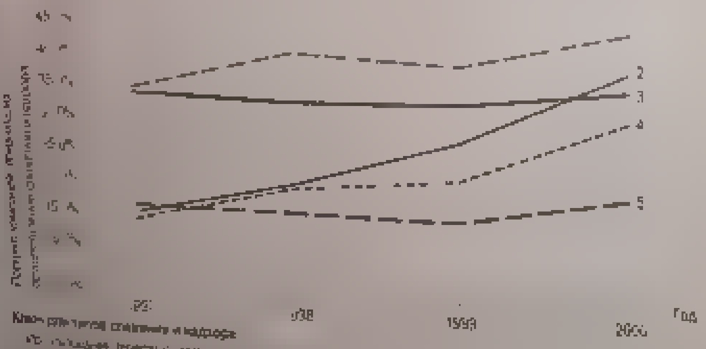


Рис. 2.9. [illegible] 1997–2000

В наш век информации использование компьютеров на рабочих местах становится все более повсеместным, но эти данные показывают, что и потенциал для слежения и надзора за рабочим местом тоже растет. Компьютер на вашем рабочем столе — это палка с двумя концами. Он предоставляет вам инструменты, необходимые для работы, а работодатели все чаще используют примерно ту же технологию для того, чтобы за вами следить.

Заключение

1. Мы рассмотрели несколько различных путей подведения итогов по распределению одной переменной и, в целом, описания результатов ваших исследований. На протяжении всей главы наш акцент был на необходимости четко и сжато документировать результаты. Часто вы будете видеть, что по мере того, как вы пытаетесь донести статистическую информацию до других, значение информации становится понятнее и для вас.

2. Процентные показатели и пропорции, отношения и удельные показатели представляют собой различные методы повышения прозрачности данных путем выражения результатов с точки зрения относительной частоты. Процентные показатели и пропорции характеризуют относительную встречаемость некоторой категории переменной в сравнении со всем распределением в целом. Отношения сравнивают две категории друг с другом, а удельные показатели характеризуют действительное число появлений некоторого события в сравнении с количеством возможных событий за некоторую единицу времени.

3. Частотные распределения представляют собой таблицы, в которых обобщается распределение значений одной переменной в целом. Очень часто построение таких таблиц для каждой переменной, представляющей интерес, является первым шагом в статистическом анализе. Столбцы, содержащие доли в процентах, кумулятивные частоты и/или кумулятивные процентные показатели, часто повышают читабельность частотных распределений.

4. Секторные и столбчатые диаграммы, гистограммы и линейчатые диаграммы (частотные [illegible]) представляют собой графические средства [illegible] выражения [illegible] эффективной форме [illegible] информации, которую содержит в себе частотное распределение.

Сводка формул

Пропорция	2	[illegible]
Процентный показатель	2.2	$\% = \left(\frac{f}{N}\right) \times 100$
Отношение	[illegible]	[illegible]

Словарь

Гистограмма. Графическое средство вывода интервальных/относительных переменных. [illegible] представлены смежными столбцами [illegible] расположенными между границами интервалов группирования. Высота каждого столбца соответствует количеству (или процентному содержанию) наблюдений в данной категории.

Интервалы группирования. Категории, используемые в частотных распределениях интервальных/относительных переменных.

Кумулятивная (накопленная) частота. Необязательный столбец в частотном распределении, указывающий количество наблюдений внутри данного интервала и во всех предыдущих интервалах.

Кумулятивный (накопленный) процентный показатель. Необязательный столбец в частотном распределении, содержащий процентный показатель наблюдений внутри данного интервала и во всех предыдущих интервалах.

Линейчатая диаграмма. См. **Частотный полигон.**

Отношение. Количество наблюдений в одной категории, деленное на количество наблюдений в некоторой другой категории.

Пропорция. Количество наблюдений в одной категории переменной, деленное на количество наблюдений во всех категориях переменной.

Процентный показатель. Количество наблюдений в категории переменной, деленное на число наблюдений во всех категориях переменной, умноженное на 100.

Секторная диаграмма. Графическое средство вывода, предназначенное для дискретных переменных, имеющих только несколько категорий. Круг (пирог) делится на сектора, размеры которых пропорциональны доле (в процентах) наблюдений в каждой категории переменной.

Столбчатая диаграмма. Графическое средство вывода дискретных переменных. Категории представляются столбцами одинаковой ширины, высота каждого из которых соответствует количеству (или процентному показателю) наблюдений в данной категории.

Удельный показатель. Количество действительных появлений некоторого явления, деленное на количество возможных появлений за некоторую единицу времени.

Центральная точка. Точка, расположенная точно посередине между верхним и нижним пределами интервала группирования.

Частотное распределение. [illegible] количество наблюдений в каждой категории переменной.

Частотный полигон. Графическое средство вывода [illegible] переменных. Интервалы группирования [illegible] [illegible] над центральными точками. Вертикальное положение каждой [illegible] соответствует количеству (или процентному содержанию) наблюдений в [illegible] интервале. Все точки соединены прямыми линиями. То же, что линейчатая диаграмма.

Мультимедийные ресурсы

[illegible] Центра социологических ресурсов Wadsworth

[illegible]

Для [illegible] студенческих ресурсов посетите web-сайт [illegible] доступа к следующим инструментам исследования:

- [illegible]
- [illegible] карточки
- [illegible]
- [illegible]
- Интернет-ссылки
- Таблица случайных чисел
- Примеры и упражнения с применением MicroCase и SPSS
- Функциональные диаграммы "Find the test"
- Проверка гипотез для переменных, измеряемых по порядковой шкале

Задачи

2.1 SOC

В следующих таблицах указано семейное положение 20 респондентов, проживающих в двух различных жилых комплексах. *Подсказка: прежде чем приступать к решению следующих задач, убедитесь, что у вас правильные числа в числителе и знаменателе. Например, в задаче 2.1а спрашивается "процентный показатель [illegible] респондентов в каждом комплексе", и знаменателем для [illegible] двух дробей будет 20. С другой стороны, в задаче 2.1г спрашивается "процентный показатель одиноких респондентов, проживающих в комплексе B", и знаменателем [illegible] будет 4 + 6, т.е. 10.*

Семейное положение	Комплекс A	Комплекс B
В браке	5	10
Не в браке [illegible]	8	2
Одинокие	4	6
Разведенные	2	1
Овдовевшие	0	1
[illegible]	[illegible]	[illegible]

а. [illegible] в браке?

б. [illegible]

з. Чему равна пропорция опрошенных в каждой выборке?

и. Чему равен процент [illegible], проживающих в комплексе В?

к. Чему равно отношение между не в браке/живущих вместе и состоящих в бра[ке] [illegible] комплексе?

2.2 [illegible]

Специальность	Мужчины	Женщины
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
Экономика	[illegible]	[illegible]
Уход за больными	3	35
[illegible]	[illegible]	5

[illegible] ответа, внимательно про[чи]тайте каждую из следующих задач

а. Чему равен процентный показатель мужчин среди всех студентов, учащих[ся] на социологических специальностях?

б. Чему равна пропорция женщин среди всех студентов, учащихся на экономических специальностях?

в. Каково отношение мужчин и женщин, учащихся на гуманитарных специальностях?

г. Какой процент общего числа студентов составляют мужчины?

д. Чему равно отношение мужчин и женщин для всей выборки в целом?

е. Чему равна пропорция мужчин среди всех студентов, учащихся на специальности "Уход за больными"?

ж. Какой процент выборки обучается на социологических специальностях?

з. [illegible] [гуманита]рных специальностях, и количеством студентов, обучающихся на экономи[ческих специальностях]?

и. [Како]ва отношение между количеством женщин, учащихся на экономических специальностях, и количеством [illegible]?

к. [illegible]

2.3 СJ

В [illegible] Канзас, с населением 21 732 человека, за прошед[ший год] [illegible] 41 ограблений банков, 13 убийств, 23 кражи автомобилей. [illegible] каждого вида преступлений на 100000 человек [illegible] ы указали численность [illegible]

2.4. CJ

Ниже приведены показатели количества уголовных преступлений, о которых было сообщено в полицию, за прошлый год по двум разным городам. Подсчитайте значение удельного показателя каждого вида преступлений на 100000 человек населения? В каком городе проблема преступности более очевидна (с относительной точки зрения)? К каким видам преступлений это относится? Напишите один абзац, описывающий различия.

Преступление	Город А (Население = 20109)	Город В (Население = 764213)
[illegible]	3	5
Ограбление	42	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	2	9
[illegible] кража со взломом	178	[illegible]

2.5 SOC

Ниже представлены [illegible] четырех [illegible] и значения [illegible] Общим общественным опросом (General Social Survey, GSS). Этот набор данных более подробно описан в приложении Ж. В нашем материале он используется для компьютерных упражнений. Небольшие подвыборки из GSS будут использоваться в материале книги для того, чтобы предоставить "реальные" данные для задач. Относительно задававшихся вопросов и прочих подробностей, см. приложение Ж. Ниже даны числовые коды для переменных.

Пол	Поддержка контроля за ношением оружия	Уровень образования	Возраст
1 = Мужской	1 = За	0 = Ниже, чем среднее	В годах
2 = Женский	2 = Против	1 = Среднее	
		2 = [illegible]	
		3 = [illegible]	
		4 = [illegible]	

Наблюдение №	Пол	Поддержка контроля за ношением оружия	Уровень образования	Возраст
[illegible]	[illegible]	1	[illegible]	45
[illegible]	2	2	[illegible]	48
3	1	[illegible]	[illegible]	55
4	2	[illegible]	[illegible]	32
[illegible]	1	1	3	33
5	2	1	[illegible]	28
6	1	1	0	[illegible]
7	2	2	[illegible]	[illegible]
8	1	1	[illegible]	[illegible]
9	1	2	[illegible]	48
10	2	1	[illegible]	[illegible]
[illegible]	1	1	[illegible]	[illegible]

Участник №	Пол	Поддержка контроля за ношением оружия	Уровень образования	Возраст
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	2
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

а. П[illegible] частотное распределение для каждой переменной. Включ[illegible] содержащий процентные показатели.

б. Постройте секторную и столбчатую диаграммы для отображения распре[illegible] поддержки контроля за ношением оружия и уровня образования.

2.6 [illegible]

[illegible]тво по делам молодежи начало программу сексуального об[illegible]ершеннолетних девушек, попадавших в суды по делам не[illegible]. Девушкам были розданы тесты, состоящие из 20 пунктов. [illegible] и до участия в программе и после завершения програм[illegible]ы значения по тестированию для первых 15 девушек, [illegible]

Наблюдение	Предварительный тест	Тест после программы
[illegible]	[illegible]	2
[illegible]	[illegible]	3
[illegible]	[illegible]	2
[illegible]	[illegible]	9
[illegible]	[illegible]	8
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	12
[illegible]	[illegible]	11
[illegible]	[illegible]	7
[illegible]	[illegible]	12
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	5
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	20

[illegible]

2.7 SOC

[illegible] для того, чтобы подготовить их к работе в советах колледжей. Их оценки представлены ниже:

420	345	560	650
459	499	500	657
467	480	505	555
480	520	530	589

Для того чтобы оценить готовность этих школьников к работе на уровне колледжа, им раздали тесты по математике и вербальным способностям. Представленные ниже оценки указывают процент правильных ответов по каждому тесту.

Тест по математике

[illegible]	[illegible]	[illegible]	[illegible]
72	85	90	99
50	73	77	78
[illegible]	[illegible]	[illegible]	[illegible]

Тест вербальных способностей

[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	70	56	60
[illegible]	78	80	92
[illegible]	[illegible]	[illegible]	[illegible]

а. Для каждой переменной постройте частотное распределение со столбцами процентных показателей и кумулятивных процентных показателей.

б. Постройте гистограмму и частотный полигон для тех же данных.

2.8 GER

Числа, указанные ниже, представляют собой число раз, которые 25 жителей комплекса для престарелых выходили на прошлой неделе из дому по какой-либо причине.

[illegible]

а. [illegible]

б. [illegible]

в. [illegible]

г. [illegible] и частотный полигон для наглядного представле[illegible]
[illegible]

д. [illegible] подводятся итоги по распределе-
[illegible]

2.9. SOC

[illegible] студентов заполнили по анкете, которая измеряла их [illegible] [illegible], которые [illegible] во многих ситуациях допустимо использовать физическ[illegible] [illegible] [illegible] ситуаций (или их очень мало), в которых можно было бы оп[illegible] [illegible]ение насилия.

[illegible]	[illegible]	1	8	92
[illegible]	[illegible]	[illegible]	9	[illegible]
17	63	[illegible]	17	23
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	25	17

а. [illegible]
б. [illegible]
в. [illegible] к таблице столбцы, содержащие процентный показатель, кумуля[illegible] частоты и кумулятивный процентный показатель.
г. [illegible] гистограмму и частотный полигон для этих данных.
д. [illegible] один абзац текста, в котором подведите итоги, касающиеся рас[illegible] значений.

2.10. PA, CJ

[illegible] эффективности работы местного отделения полиции были собра[illegible] данные о времени реагирования полиции на звонки о помощи [illegible] время реагирования было округлено до целых минут). Переведите [illegible] распределения в процентные показатели и постройте секторные [illegible] диаграммы для этих данных. Напишите абзац, в котором прове[illegible] времени реагирования между этими двумя годами.

Время реагирования (мин), 1986	f	Время реагирования (мин), 1996	f
[illegible]	75	21 и более	45
[illegible]	[illegible]	6–20	75
[illegible]	60	15	66
[illegible]	5	6–[illegible]	35.0
[illegible]	210	[illegible]	[illegible]

2.11. SOC

[illegible]

менение удельных показателей преступности происходит в одном [illegible] ник? Обратите внимание на то, как различаются вертикальные оси [illegible]. Для убийств ось имеет значения от 0 до 12, тогда как для ночных краж с [illegible] и краж автомобилей — от 0 до 1600. Последние из названных преступлений [illegible] шкала с меньшими интервалами.

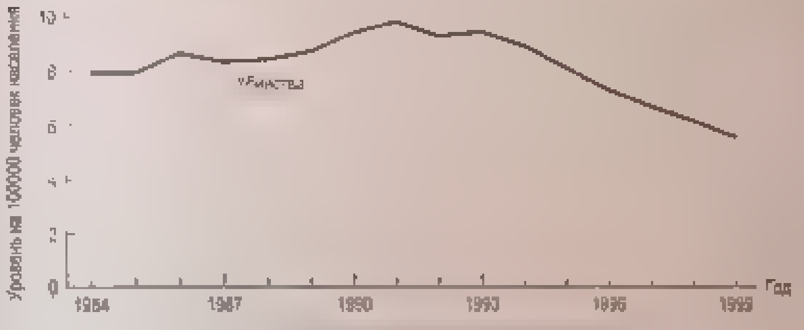


Рис. 2.10. [illegible]

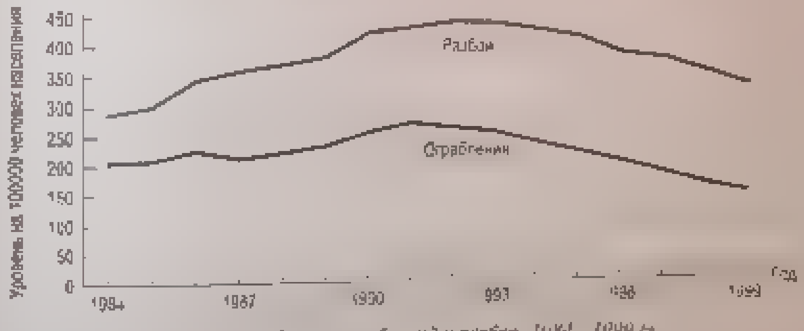


Рис. 2.11. *Уровни ограблений и разбоя, 1984—1999 гг.*

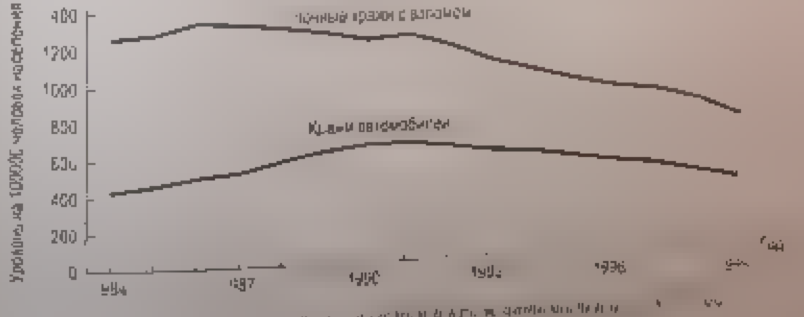


Рис. 2.12. [illegible]

2.12. [illegible] количества дорожно-[illegible] [illegible] полицейский патруль. Данные о [illegible] понижен предел скорости и усилен полицейский патруль. Данные о количестве [illegible] представлены ниже. Какое влияние оказа[illegible] [illegible] программное обеспечение, если оно вам доступно) и напишите [illegible]

	12 месяцев до	12 месяцев после
[illegible]	[illegible]	[illegible]
Февраль	25	21
Март	20	18
Апрель	19	12
Май	15	9
Июнь	17	10
Июль	24	11
Август	28	15
Сентябрь	23	17
Октябрь	[illegible]	[illegible]
Ноябрь	21	[illegible]8
Декабрь	22	20

[illegible]

Создание частотных распределений и графиков с помощью *SPSS for Windows*

[illegible] SPSS [illegible] щелкните по значку SPSS на экране своего монитора. Загрузите 1998 GSS, щелкнув [illegible] файл в первом окне или нажав File (Файл), Open (Открыть) и Data (Данные) в окне SPSS Data Editor [illegible] SPSS [illegible] GSS98.sav [illegible] сделайте двойной щелчок кнопкой мыши по имени файла, чтобы [illegible] данные. Когда вы увидите [illegible] SPSS [illegible] цессор SPSS готов"), это означает, что можно приступать к работе.

Демонстрационный пример с использованием пакета SPSS 2.1. Частотные распределения

[illegible] Analyze [illegible] Descriptive Statistics [illegible] Frequencies [illegible] Frequencies [illegible]

[illegible] MAN WANTS [illegible] [illegible] SPSS [illegible] переменных, найдите в левом окне переменную *marital*. [illegible] Variables [illegible] SPSS [illegible] OK [illegible] SPSS [illegible] Output [illegible] Output [illegible] [illegible]атов).

[illegible] *marital* [illegible]

Семейное положение

	Частота	Процент	Действительный процент	Кумулятивный процент
Действительные				
В браке	650	46,9	46,9	46,9
Овдовевшие	134	9,7	9,7	56,6
Разведенные	216	15,6	15,6	72,2
Разъехавшиеся	51	3,7	3,7	75,8
Не состоявшие в браке	335	24,2	24,2	100,0
[illegible]	[illegible]	[illegible]	[illegible]	
Отсутствует	1	0,1		
Сумма	1387	100,0		

[illegible] MAR[illegible] [illegible] один вправо) находится действительные значения частот, [illegible] ко возникало каждое из значений переменной. Мы видим [illegible] состоят в браке, 134 овдовели, и т.д. В следующих [illegible]

[illegible] дали) даны из расчета на всех ре[illegible]

[illegible]

и т. д.), тогда число категорий можно снизить до 8. Можно сократить число к[illegible] [illegible] демонстра[illegible] примере с использованием пакета SPSS 2.2

Демонстрационный пример с использованием пакета SPSS 2.2
Сокращение числа категорий с помощью команды Recode

Часто для того, чтобы создать читабельные частотные распределения, нужно бу[illegible] сокращать количество значений интервальн[illegible] SPSS for Windows [illegible] Recode [illegible]

[illegible] соответствовать 10 годам. Так как самым [illegible]

дующим интервалом будет 20–29, затем 30–39 и так далее до интервала 80–89. Как вы увидите далее, перекодирование включает в себя массу маленьких шагов, поэтому прис[illegible] запастись терпением и поочередно выполнить команды.

1. [illegible] SPSS Data Editor [illegible] Recode [illegible] Same Variable [illegible] Different Variable [illegible]

2. [illegible] Recode into Different Variable [illegible] Input Variable → Output Variable [illegible]

3. В поле Output Variable (Выходная переменная), расположенном справа, щелкните по полю Name (Имя) и введите имя для новой (выходной) переменной. Я предлагаю назвать новую переменную *ager* (*age recoded*, age перекодированная), но вы можете назвать ее любым именем, если только оно не повторяет имя другой переменной, хранящейся в наборе данных, и его длина не превышает восьми символов. Нажмите кнопку **Change** (Изменить), и в поле **Input Variable → Output Variable** (Входная переменная → Выходная переменная) появится выражение *age → ager*.

4. Нажмите на кнопку **Old and New Values** (Старые и новые значения) в центре экрана, и откроется новое диалоговое окно. Найдите в левом столбце кнопку **Range** (Диапазон). Нажмите на эту кнопку, и курсор переместится в маленькое поле, находящееся прямо под ней. В этих полях следует указать наименьшую и наибольшую точку каждого интервала новой переменной *ager*.

5. Начнем с интервала 10–19. Введите в расположенное слева диалоговое окно **Range** (Диапазон) значение 10, после чего щелкните по полю, расположенному справа, и введите 19. В поле **New Value** (Новое значение), расположенном в верхнем [illegible] **Value** [illegible] **Value** введите 1, после чего нажмите кнопку **Add** (Добавить), расположенную прямо под ним. В диалоговом окне **Old → New** (Старые → Новые) появится выражение 10–19 → 1. Это является завершением первой инструкции по перекодированию для SPSS.

6. [illegible] **Range** [illegible] **New Values** [illegible] **Value** [illegible] **Add** [illegible]

бошты. В диалоговом окне Old → New (Старые → Новые) [illegible]

[illegible] New [illegible] кнопку Continue (Продолжить) [illegible] Recode into Different Variables [illegible] OK [illegible] SPSS [illegible]

[illegible] которое вы присвоили записанной переменной. SPSS добавляет новую переменную в набор данных, и вы сможете найти ее в последнем столбце справа от окна данных. Вы можете сделать новую переменную постоянной частью набора данных, сохранив в конце сеанса [illegible] нужно сохранить новый, расширенный файл данных, нажмите No (Нет) в ответ на вопрос, хотите ли вы сохранить файл данных. Если вы [illegible] SPSS для Windows, помните, что вы ограничены [illegible]

Чтобы создать для новой переменной [illegible] распределение, нажмите Analyze [illegible] Descriptive Statistics [illegible] Frequencies (Частоты). В поле, расположенном слева, найдите переменную [illegible] кнопке-стрелке, чтобы переместить ее в поле, расположенное справа. Щелкните OK [illegible]

AGER — сокращенная версия AGE

		Частота	Процент	Действительный процент	Кумулятивный процент
[illegible]					
Сумма		1387	100,0		

[illegible]

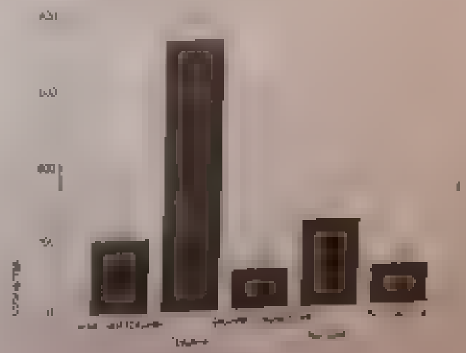
Демонстрационный пример с использованием SPSS 2.3. Графики и диаграммы

[illegible] я сведу количество опций к минимуму, но вам нужно исследовать их самостоятельно [illegible] Help [illegible]

[illegible] Graphs [illegible] Bar [illegible] Chart [illegible] Simple [illegible] му просто нажмем Define (Определить). Появится диалоговое окно Define Simple Bar (Определить простую столбчатую диаграмму), в левой части которого будут [illegible] ны имена переменных. Из списка переменных выберите *degree* (*степень*), переместив курсор так, чтобы он выделил имя этой переменной. Нажмите на кнопку-стрелку [illegible] переместить переменную *degree* в поле Category Axis (Ось категорий).

[illegible] внимание на то, что поле Bars Represent (Столбцы представляют) расположено выше, чем поле Category Axis (Ось категорий). Это [illegible] на которую можно [illegible] давайте выберем уже выделенную опцию N of cases or frequencies (Количество наблюдений или частоты).

[illegible] Options [illegible] Options [illegible] Display groups defined by missing values (Отображать группы, образованные пропущенными значениями) [illegible] выделена (или отмечена). Нажмите кнопку Continue (Продолжить), после чего нажмите кнопку OK в диалоговом окне Define Simple Bar (Определить простую столбчатую диаграмму). В результате должна получиться следующая столбчатая диаграмма.



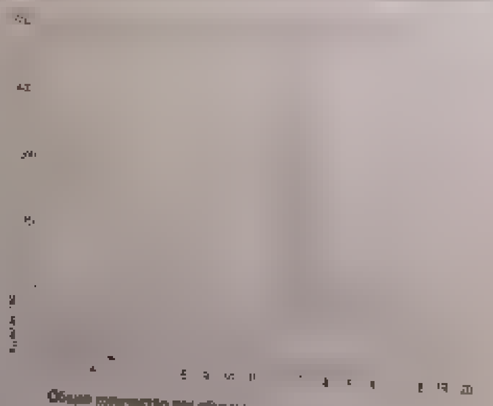
Высшая из полученных степеней

[illegible] SPSS Chart Editor [illegible] по любой части графика [illegible]

[illegible] Format [illegible]

[illegible] Graphs [illegible] Line [illegible] Line Charts [illegible] Simple (Прост[illegible] нам и нужна, поэтому мы нажмем Define (Определить), после чего появится диало[illegible] Define Simple Line [illegible] ваши действия должны быть такими же, как и в диалоговом окне Define Simple Bar (Определить простую столбчатую диаграмму). Линейчатые диаграммы подходят [illegible] интервальных/относительных данных, поэтому используем для демонстрации переменную *educ*. Эта переменная представляет собой своего рода несокращенную форму переменной *degree*, и мы сможем быстро и легко сравнить две переменные.

Из списка переменных выберите *educ* и нажмите на кнопку-стрелку в середине экрана, чтобы переместить *educ* в поле Category Axis (Ось категорий). Обратите внимание на то, что диалоговое окно Lines Represent (Линии представляют) такое же, как и диалоговое окно Simple Bar (Простая столбчатая диаграмма). Нам нужно, чтобы столбцы представляли *частоты*, и эта опция уже выбрана. Нажмите кнопку Options (Опции) и убедитесь, что опция Display groups defined by missing values (Ото[illegible] Continue [illegible] OK [illegible] Define Simple Line (Определить простую [illegible] диаграмму [illegible]



Линейчатую диаграмму можно редактировать, щелкнув на любой части диаграммы и вызвав окно Chart Editor (Редактор диаграмм). Например, для того чтобы из[illegible] линий, щелкните по какой-либо линии в окне Chart Editor (Редактор диаграмм), [illegible] чего нажмите Format (Формат) и Color (Цвет). Выберите новый цвет [illegible] Apply [illegible] Save [illegible] Print [illegible]

[illegible] ненную при построении столбчатой диаграммы для переменной *degree*

Упражнения

2.1. [illegible] вых переменных, включая *race* (*расовая принадлежность*) и *relig* (*религиозная принадлежность*). Постройте столбчатые диаграммы для каждой переменной и напишите одно-два предложения по поводу каждого частотного распределения. Ваше описание должно ясно показывать наиболее и наименее часто встречающиеся значения и любые други[illegible], достойные внимания.

2.2. Постройте частотные распределения для переменных *degree* и *educ*. Используйте команду Recode (Перекодировать, SPSS) или Collapse Variables (Свернуть переменные, MicroCase) для того чтобы свести переменную *educ* до числа категорий, сравнимых с *degree*. То есть, используйте категории 0–[illegible] для уровня образования ниже, чем среднее образование, 12 для выпускников средних школ, 13–15 для выпускников технических колледжей, 16 для выпускников университетов, 17–20 для аспирантов. Постройте частотное распределение для перекодированной переменной *educ* и сравните ее с частотным распределением для переменной *degree*. Опишите три частотных распределения в нескольких [illegible]

2.3. Постройте частотное распределение для переменной *prestg80* (престиж должности, занимаемой респондентом). Найдите диапазон изменения этой переменной и, используя данную информацию, определите разумные интервалы группирования, следуя процедурам, описанным в разделе 2.5. С помощью команды Recode (Перекодировать, SPSS) или Collapse Variables (Свернуть переменные, MicroCase), создайте итоговое частотное распределение для перекодированной переменной и напишите одно-два предложения с описанием и интерпретаци[illegible]

2.4. Выберите еще пять-шесть интересующих вас переменных и постройте столбча[illegible] столбчатые диаграммы [illegible] менных, и линейчатые диаграммы — для интервальных/относительных переменных [illegible] рамм.

3 Показатели центра распределения

Цели изучения

К концу этой главы вы сможете

1. Объяснять назначение показателей центра распред[illegible] трактовать заключенную в них информацию
2. Рассчитывать, объяснять, сравнивать и со[illegible] статистические показатели, как мода, [illegible] арифметическое
3. [illegible] смысл средне[illegible]
4. Выбирать подходящий показатель [illegible] соответствии со шкалой измерения и асимметрией

3.1. Введение

Одно из очевидных преимуществ использования показателя распределения час тот, диаграмм и графиков заключается в том, что они обобщают форму распределения значений таким образом, что их можно легко понять. Тем не менее вам часто будет необходимо представить более подробную информацию о распределении. Почти всегда полезными являются два вида статистической информации: понятие о типичном или среднем наблюдении, взятом из распределения (например, средняя [illegible] зарплата социальных работников составляет от $19 000 до $28 000 в год). Первый вид [illegible] показателями центра распреде[illegible] данной главы. Второй вид данных — показатели изменчивости [illegible] будут рас[illegible]

Вам должны быть знакомы три наиболее час[illegible] [illegible] арифметическое. Все эти показатели обоб[illegible]

[illegible]

[illegible] чительно отличаются друг от друга. В действительности, одинаковое значение они [illegible] меньше, и что, пожалуй, более существенно как они отличаются при определении центра распределения — они не обязательно будут определять одно и то же значение или наблюдение как "типичное". Таким образом, ваш выбор подходящего показателя центра распределения будет зависеть частично от способа, которым вы измеряете переменные, и частично, от цели исследования.

3.2. Мода

Мода распределения — это значение, которое появляется чаще всего. Например, в ряду значений 58, 82, 82, 90, 98 мода будет равняться 82, поскольку это значение появляется два раза, в то время как другие значения попадаются лишь один раз.

В общем мода является довольно простым показателем, наиболее полезным в том случае, когда вам необходим "простой и быстрый" показатель центра распределения и когда вы работаете с номинальными переменными. Фактически, мода является единственным показателем центра распределения, который может быть применен для переменных, измеряемых по номинальной шкале. Такие показатели, конечно же, не имеют числового значения, и модой номинальной переменной будет ее наибольшая категория. Например, в таблице 3.1 показана религиозная принадлежность вымышленной выборки из 242 респондентов. Мода распределения — это наибольшая категория, т.е. протестанты.

Таблица 3.1. Религиозная принадлежность (вымышленные данные)

Вероисповедание	Частота
[illegible]	[illegible]
Католик	57
Иудей	10
Атеист	32
[illegible]	[illegible]
	N = [illegible]

Если исследователь хочет показать лишь наиболее распространенное значение в распределении или если рассматриваемая переменная является номинальной, то мода является подходящим показателем центра распределения. Тем не менее помни[illegible] [illegible] вообще не имеют моды (см. таблицу 2.5) или имеют столько мод, что это теряет [illegible]

[illegible]

[illegible] Например [illegible] [illegible] распределение баллов теста по ста[illegible]

[illegible]

Баллы [illegible]	Частота
[illegible]	[illegible]
N	[illegible]

[illegible] распределения равна 93. Является ли это значение наиболее близким к большинству значений? Если преподаватель обобщит данное распределение, показав лишь значение моды, даст ли это общее представление о распределении в целом? (*Для развития навыков определения и пояснения моды, см. задачи 3.1—3.7*)

3.3. Медиана

В отличие от моды, медиана (*Md*) всегда точно представляет центр значений распределения. Медиана — это значение наблюдения, находящегося точно в середине распределения: половина наблюдений имеет значения больше, а половина наблюдений — меньше, чем наблюдение, характеризующееся серединным значением. Таким образом, если медиана семейных доходов равна $30 000, то половина семей зарабатывает больше $30 000, а половина — меньше.

Прежде чем определять медиану, наблюдения должны быть упорядочены от большего до меньшего значения, либо от меньшего к большему. Как только это выполнено, находится центральное наблюдение. Медиана — это значение такого наблюдения. Если пять студентов получило по тесту 93, 87, 80, 75 и 61 балл, то медиана [illegible]

Когда количество наблюдений (*N*) является нечетным, значение медианы однозначно [illegible] В то же время, [illegible] ство наблюдений четно, серединных значений будет два. В таком случае [illegible] ределяется как значение, [illegible]

[illegible] Поскольку [illegible] будет четвертое наблюдение, [illegible] выше и три наблюдения ниже него. Если семь значе-

[illegible] было бы значение (25+1)/2 или тринадцатого наблюдения.

Теперь, если мы добави[illegible] ние значения теперь будет иметь вид 10, 10, 8, 7, 5, 4, 2, 1, в общем то, любое [illegible] ит распределение на две равные половины по четыре наблюдения каждая). Такая неопределенность разрешается путем определения медианы как среднего значения двух серединных наблюдений. В приведенном выше примере медиана будет равна [illegible]

Для определения двух серединных значений, когда *N* равно четному числу, чтобы найти первое серединное наблюдение разделите *N* на 2 [illegible], для того чтобы найти второе серединное на[illegible]ьте полученное значение на 1. В приведенном примере восемь наблюдений, первое серединное наблюдение будет четвертым наблюдением ($N/2 = 4$), а второе наблюдение будет $(N/2) + 1$ или пятым наблюдением. Если бы *N* равнялось 142, первым серединным наблюдением было бы 71-е, а вторым — 72-е наблюдение. Помните, что медиана определяется как среднее из значений, связанных с двумя серединными наблюдениями.

Поскольку медиана требует, чтобы значения были упорядочены от большего к меньшему, она не может быть определена для переменных, измеренных по номинальной шкале. Помните, что значения переменных, измеряемых по номинальной шкале измерений, не могут быть упорядочены: значения отличаются друг от друга, но не дают оснований проводить математическое сравнение. Медиана может быть определена для порядковых либо для интервальных данных, но в целом наиболее подходит для первого типа (для любой из задач в конце данной главы медиану рассчитать можно.)

3.4. Прочие показатели положения: процентиль, дециль, квартиль

Кроме того, что медиана является показателем центра распределения, она также [illegible] ют распределение на три или четыре части, или точки, ниже которых [illegible]

[1] Если серединные наблюдения имеют одинаковое значение, [illegible] значения равные 6, и, таким образом, медиана [illegible]

[illegible]

по нормированным тестам, которые часто позволяют и распред...
... равное 476 баллам выше, чем 46% всех значений)

... является часто используемым ...
... Процентиль показывает точку, ниже которой находится опре...

процентили. Для того, чтобы найти 37-ю процентиль для выборки из 78 наблюд... нии, умножьте 78 на 0,37. Результат равен 28,86, а 37-я процентиль будет 86/100 рас... стояния между значениями 28-го и 29-го наблюдений. В большинстве случаев мы округлим 28,86 до 29 и назовем значение 29-го наблюдения 37-ой процентилью. Не... большая неточность будет компенсирована экономией времени и сокращением рас...

Обратите внимание, что если мы рассматриваем процентиль, то медиана будет просто 50-й процентилью и, в приведенном выше примере, мы найдем медиану умножив 78 на 0,50, получим 39-е наблюдение и скажем, что значение этого наблюдения является 50-й процентилью. Еще раз обратите внимание, что мы в данном случае немного упрощаем. Фактически медианой будет значение посредине между двумя серединными наблюдениями (39-м и 40-м), но вряд ли эта неточность будет существенной.

Другими часто используемыми показателями положения являются децили и квартили. Дециль делит распределение на значения по десяткам процентов. Таким образом, первый дециль является точкой, ниже которой находятся 10% наблюдений и она равна 10-й процентили. Пятая дециль также равна 50-ой процентили, что в свою очередь равно (вы уже догадались) медиане. Квартили делят распределение на четыре...
... а третья квартиль ...

3.5 Среднее арифметическое

Среднее арифметическое ... — читается как "икс с чертой") является наиболее ча... ... распределения. Оно показывает среднее зна...

... распределения и рассчитывается следующим образом: ... арифметического сложите все значения и разделите полученную сумму на количество значений N. Например, ...

результатом по тесту.

Математическая формула среднего арифметического имеет вид

Формула 3.1 $$\bar{X} = \frac{\sum X_i}{N}$$

$\bar{X}$ — среднее арифметическое,

$\sum(X_i)$ — сумма значений,

N — количество значений.

Поскольку в данной формуле представлены некоторые новые символы, давайте рассмотрим их. Во-первых, символ $\sum$ (прописная греческая буква сигма) является математическим оператором таким же, как знак плюс (+) или знак деления (÷). Этот оператор обозначает "сумму", то есть мы должны сложить любое количество чисел, находящихся за данным знаком. Второй новый символ — это X_i ("X i-тое"), он может обозначать любое отдельное значение — i-тое значение. Если мы хотим выделить определенное значение в распределении, определенное число должно заменить нижний индекс. Таким образом, X_1 будет обозначать первое значение, X_2 — второе, ... — двадцать шестое и т.д. Операция суммирования всех значений будет обозначаться как $\sum(X_i)$.

Данная комбинация символов показывает, что необходимо суммировать значения, начиная с первого, заканчивая последним значением распределения. Таким образом, формула 3.1 показывает в символах то, что уже было описано (для расчета среднего арифметического сложите значения и разделите на количество значений), но очень кратко и точно[4] (для тренировки расчета среднего арифметического, см. любую из задач в конце данной главы).

Поскольку для расчета среднего необходимо выполнять операции сложения и деления, ее использование полностью оправдано только при работе с интервальными ... относящихся и к порядковой шкале, поскольку среднее арифметическое ... гораздо более гибким показателем, чем медиана, и оно является основой ... для многих интересных и сложных статистических расчетов. Таким о... среднее будет, пожалуй, подходящим показателем центра ра...

3.6. Некоторые свойства среднего арифметического

[illegible] более часто используемым по[illegible] [illegible] математические и статистические [illegible] [illegible] арифметическое всегда является центром любого рас[illegible] [illegible] [illegible]

$$\sum [illegible] \overline{X} = [illegible]$$

[illegible] в распределении [illegible] тем среднее арифмети[illegible] [illegible] результатом такого суммирования всегда будет ноль. Например, рассмотрим следующий набор баллов по тесту: 65, 7[illegible] [illegible] этих значений [illegible] 390/5 или 78. [illegible]

X	$X - \overline{X}$
65	65 − 78 = [illegible]
7[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
$\sum$ [illegible]	$\sum (X - \overline{X}) =$ [illegible]

Сумма отрицательных разностей (−19) точно равна сумме положительных разностей (+19), и эти значения всегда будут равны вне зависимости от распределения [illegible] [illegible] зывают, что среднее арифметическое является хорошим описательным показателем [illegible] значений. Вы можете считать среднее арифметическое точкой, которая точно уравновешивает все значения.

[illegible] свойство среднего арифметического на данном этапе мы лишь упоминаем, [illegible] первостепенная важность проявится в последующих главах. Данное свой[illegible] [illegible] как:

$$\sum (X - \overline{X})^2 = \text{минимум}$$

[illegible footnotes]

[illegible]

регрессии, которые мы рассмотрим позднее.

Последнее важное свойство среднего арифметического состоит в том, что любое значение в распределении влияет на него. Мода (которая представляет собой всего [illegible] лишь значением срединных наблюдений (или наблюдения) не столь подвержены влиянию. Данное свойство является и преимуществом, и недостатком. С одной стороны, среднее арифметическое использует всю имеющуюся информацию — любое значение в распределении влияет на среднее. С другой стороны, когда в распределении несколько экстремальных наблюдений (значения слишком большие или малые), среднее арифметическое может стать показателем центра распределения, вводящим в заблуждение.

Например, рассмотрим следующий набор значений: 15, 20, 25, 30, 35. Среднее арифметическое и медиана данного распределения равны 25. ($\overline{X} = 125/5 = 25$, Md = значение третьего наблюдения = 25). Что произойдет, если мы изменим последнее значение с 35 до 3500? Обратите внимание, что с медианой ничего не произойдет. Она по-прежнему будет равняться 25. Медиана основывается лишь на значении серединного наблюдения и не подвергалась влиянию изменений значений других наблюдений в распределении. Среднее же, поскольку оно учитывает все значения, сильно изменилось. Оно стало равно 3590/5, или 718. Очевидно, что одно экстремальное значение в наборе данных непропорционально влияет на среднюю арифметическое. В таком случае, какой из двух показателей более точно отражает центр данного распределения (*для отработки понимания влияния экстремальных значений на средние арифметические и медианы см. задачи 3.11, 3.13, 3.14 и 3.15*)?

Какой общий вывод мы можем сделать из данных рассуждений? В отличие от медианы среднее арифметическое всегда смещено в сторону экстремальных значений. Два показателя центра распределения будут иметь одинаковое значение лишь и только тогда, когда распределение симметрично. Когда в распределении содержится несколько экстремально больших значений (положительная асимметрия), среднее арифметическое всегда будет иметь больше [illegible] Если в распределении содержится несколько [illegible] среднее арифметическое [illegible] на. Рисунки 3.1—3.3 показывают три различных полигона частот, которые [illegible]рируют рассмотренные взаимосвязи.

[illegible] между [illegible] и средним арифметическим [illegible] [illegible] метического всегда покажет — является ли распределение асимметричным [illegible]

[illegible] рии. Во-вторых, они являются простыми и удобными [illegible] [illegible]. Например, если [illegible] распределения, укажите медиану. Данные [illegible]

[illegible] богатых людей. Если вы хотите поразить кого-либо данными о достатке групп [illegible] данными доходами, укажите среднее арифметическое. [illegible]

[illegible] сильно асимметричного распределения будет зависеть от того, что он или она х[illegible]

Рис. 3.1

Рис. 3.2

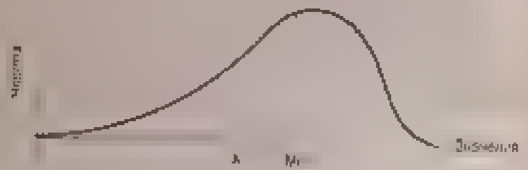
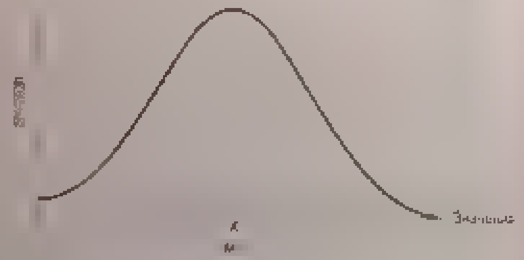


Рис. 3.3
Симметричное расп[illegible]



Практический пример 3.1

[illegible] сколько часов они провели в библиотеке [illegible] [illegible] мы рассчитаем моду, медиану и среднее арифметическое для [illegible]

Просматривая значения, мы видим, что два значения (5 и 14) появляются [illegible]

[illegible] среднее значение [illegible] расставить их по порядку. Принимая во внимание то, что мы имеем дело с 10 наблюдениями, первым серединным наблюдением будет (N/2) или (10/2) [illegible] наблюдение. Вторым серединным наблюдением будет (N/2)+1 или (10/2)+1 [illegible] шестое наблюдение. Медиана — это значение посредине между [illegible]

[illegible]

$$\bar{X} = \sum \frac{X_i}{N} = \frac{107}{10} = 10{,}7$$

Эти десять студентов провели в среднем 10,7 часов в библиотеке за рассматри[illegible]

[illegible] умеренной асимметрией распределения [illegible] рии была вызвана тем, что [illegible] студенте провели [illegible] восемь

3.7. Выбор показателя центра распределения

При выборе показателя центра распределения [illegible] рении, по которой измеряется переменная, а также исходить из того, что показыва

Корней Чуковский

[illegible]

Используйте моду, когда:

1. Переменные относятся к номинальной шкале.
2. Вам нужен простой показатель для порядковых и интервальных переменных.
3. Вам необходимо указать чаще всего встречающееся значение.

Используйте медиану, когда:

1. Переменные относятся к порядковой шкале.
2. Переменные, относящиеся к интервальной шкале, имеют сильно асимметричные [illegible]
3. Вам необходимо указать серединное значение. Медиана всегда располагается точно в центре распределения.

Используйте среднее арифметическое, когда:

1. Переменные измеряются по интервальной шкале (кроме сильно асимметричных [illegible]).
2. Вам необходимо указать наиболее ожидаемое значение. Среднее арифметическое является центром точного равновесия всех значений.
3. Вы собираетесь проводить дополнительный статистический анализ.

Заключение

1. Три показателя центра распределения [illegible] Правильное использование этих показателей [illegible] исследователю возможность представлять важную [illegible]
2. [illegible]
3. [illegible]
4. [illegible] используется среднее арифметическое ($\overline{X}$). [illegible] значение. Его использование наиболее це[illegible] (за исключением случаев, когда распреде[illegible]

5. Среднее арифметическое обладает рядом математических свойств, которые [illegible] для проведения статистического анализа. Во-первых, оно является точкой в распределении значений, вокруг которой сбалансированы все другие значения. [illegible] В отличие от моды и медианы, на среднее арифметическое влияет каждое [illegible] значение.

Формулы

Среднее 3.1 $\overline{X} = \frac{\sum(X_i)}{N}$

Словарь

Асимметрия. Показывает, что распределение значений имеет несколько слишком больших (положительная асимметрия) либо слишком малых (отрицательная асимметрия) [illegible]

Дециль. Точка, разделяющая распределение значений на десятые доли.

Квартиль. Точка, которая делит распределение на четверти.

Медиана (Md). Точка в распределении значений, до и после которой находится [illegible]

Мода. Наиболее часто встречающееся значение в распределении или наибольшая категория переменных.

Показатели центра распределения. Статистические показатели, которые обобщают распределение значений, показывая наиболее типичное или репрезентативное [illegible] распределения.

Процентиль. Точка, за которой находится определенное количество значений.

Среднее арифметическое. [illegible] среднее арифметическое выборки, а μ — среднее арифметическое генеральной совокупности.

Σ (прописная греческая буква сигма). Суммирование.

X_i ("X с индексом i"). Любое значение в распределении.

Мультимедийные ресурсы

The Wadsworth Sociology Resource Center: Virtual Society (Виртуальное сообщество Центра социологических ресурсов Вадсворта)

http://sociology.wadsworth.com

Посетите сайт шестого издания книги *Statistics: A Tool for Social Research* [illegible] доступ к различным ресурсам для [illegible] Resources (Ресурсы для студентов), для того, чтобы получить доступ к следующим учебным материалам:

- Обзор основ математики.
- Обучающие карточки.

[illegible]

- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible] и упражнения с использованием пакетов статистической обработ[illegible] данных MicroCase и SPSS
- [illegible]
- [illegible]

Задачи

3.1 [illegible]

[illegible] информация, в т.ч. о месте рождения, о том, насколько они [illegible] легализацию марихуаны (измерено по шкале, где 7 = сильная [illegible] 4 = нейтральное отношение, 1 = полное неприятие), о количестве [illegible] денег, которое они еженедельно тратят на еду, напитки и развлече[illegible] сколько фильмов они посмотрели в общей комнате за последнюю [illegible] их отношении к еде из кафетерия (10 = превосходная, 0 = очень [illegible] и об их вероисповедании. Часть [illegible]

Первокурсники

Студент	Место рождения	Легализация	Наличные деньги	Фильмы	Еда в кафетерии	Вероисповедание
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Выпускники

Студент	Место рождения	Легализация	Наличные деньги	Фильмы	Еда в кафетерии	Вероисповедание
Л	Север	7	56	0	1	Атеист
М	Средний запад	6	52	5	2	Протестант
Н	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
О	Север	5	80	3	4	Католик
П	Юг	1	52	4	3	Протестант
Р	Юг	5	47	14	6	Протестант
С	Запад	6	30	0	2	Католик
Т	Запад	7	39	7	9	Атеист
У	Север	3	35	5	4	Атеист
Ф	Запад	5	75	3	7	Другие религии
Х	Север	4	68	5	4	Атеист

3.2 Была собрана различная информация о каждой из девяти средних школ района. Определите наиболее подходящий показатель центра распределения для каждой переменной и обобщите полученную информацию. *Подсказка: шкала измерения, по которой измерены переменные, подскажет вам, какой из показателей центра распределения является наиболее подходящим. Не забудьте расставить значения от большего к меньшему, прежде чем определять медиану.*

Средняя школа	Всего человек	Наибольшая расовая/этническая группа	Количество школьных правил	Наиболее популярный вид спорта	Физическое состояние (шкала 1–10, 10 = высокое)
1	1400	Белый	25	Американский футбол	10
2	1223	Белый	77	Бейсбол	7
3	876	Черный	52	Американский футбол	5
4	1567	[illegible] американец	28	Американский футбол	8
5	778	Белый	48	Баскетбол	4
6	1690	Черный	35	Баскетбол	5
7	250	Белый	66	Футбол	6
8	970	Белый	54	Американский футбол	9
9	1109	Латиноамериканец	84	Футбол	3

3.3 [illegible]

[illegible] вом городе и собрали некоторую информацию [illegible] боржи преданных сторонников партии. [illegible] центра распре[illegible] каждой [illegible]

[illegible]	Пол	Социальный класс	Количество лет пребывания в партии	Образование	Семейное положение	Количество детей
[illegible]	[illegible]	[illegible]	[illegible]	[illegible] школа	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	Средняя школа	[illegible]	6
[illegible]	[illegible]	[illegible]	12	Средняя школа	Холост	0
[illegible]	[illegible]	[illegible]	50	Восемь классов	Вдовец	[illegible]
[illegible]	[illegible]	[illegible]	25	[illegible] классов	Жена	4
[illegible]	М	[illegible]	25	Средняя школа	Разведен	[illegible]
[illegible]	Ж	[illegible]	2	Вуз	Разведена	[illegible]
[illegible]	Ж	[illegible]	[illegible]	Вуз	Разведена	[illegible]
[illegible]	Ж	[illegible]	2	[illegible]	Замужем	[illegible]
[illegible]	[illegible]	[illegible]	35	Вуз	Замужем	[illegible]
[illegible]	М	[illegible]	[illegible]	Средняя школа	Холост	[illegible]
[illegible]	[illegible]	[illegible]	5	Средняя школа	Разведена	0
[illegible]	[illegible]	[illegible]	3	Восемь классов	Вдова	[illegible]

3.4 SOC

Вы собрали представленную ниже информацию за 10 лет о выпускниках, которые по результатам голосования, проводившегося в местной школе, имели [illegible] [illegible] показатель центра распределения.

Наблюдение	Текущий доход	Семейное положение	Наличие BMW	Количество лет обучения после средней школы
[illegible]	[illegible]	Холост, не замужем	Нет	8
[illegible]	[illegible]	Разведен, разведена	Нет	4
[illegible]	[illegible]	Женат, замужем	Да	4
[illegible]	[illegible]	Женат, замужем	Нет	4
[illegible]	[illegible]	Холост, не замужем	Нет	4
[illegible]	[illegible]	Разведен, разведена	Да	9
[illegible]	[illegible]	Женат, замужем	Нет	3
[illegible]	[illegible]	Женат, замужем	Нет	[illegible]
[illegible]	[illegible]	Женат, замужем	Да	6
[illegible]	[illegible]	Холост, не замужем	Да	4

3.5 SOC

[illegible] данные по 4 переменным для 15 респондентов. Определите и [illegible] показатель центра распределения для каждой переменной.

Респондент	Семейное положение	Этническая группа	Возраст	Отношение к аборту*
А	[illegible] замужем	Белая [illegible]	[illegible]	0
Б	[illegible] замужем	Латиноамериканец	20	9
В	[illegible]	Белая [illegible]	21	6
Г	[illegible]	Белая [illegible]	40	0

Респондент	Семейное положение	Этническая группа	Возраст	Отношение к аборту*
Д	Женат, замужем	[illegible]	25	[illegible]
Е	Женат, замужем	[illegible]	25	[illegible]
Ж	Разведен, разведена	[illegible]	[illegible]	[illegible]
З	Вдовец, вдова	[illegible]	[illegible]	[illegible]
И	Разведен, разведена	[illegible]	3[illegible]	[illegible]
К	Женат, замужем	[illegible]	5[illegible]	[illegible]
Л	Вдовец, вдова	[illegible]	[illegible]	[illegible]
М	Женат, замужем	[illegible]	[illegible]	[illegible]
Н	Разведен, разведена	Белый [illegible]	[illegible]	[illegible]
О	Женат, замужем	Белый [illegible]	[illegible]	[illegible]
П	Разведен, разведена	[illegible]	[illegible]	[illegible]

* [illegible]

3.6 SOC

Ниже следуют четыре переменные для 30 наблюдений, взятых из Общего социологического опроса. Возраст представлен в годах. Переменная счастья состоит из ответов на вопрос: "В общем вы бы сказали, что вы (1) очень счастливы (2) довольно таки счастливы или (3) не слишком счастливы". Респондентов спросили — сколько партнеров они имели за последние пять лет. Ответы были измерены по следующей шкале: 0–4 = действительное количество, 5 = 5–10 партнеров; 6 = 11–20 партнеров, 7 = 21–100 партнеров; 8 = более 100 партнеров. Для каждой переменной подберите подходящий показатель центра распределения и напишите предложение, содержащее эту статистическую информацию, как вы бы это сделали в отчете об исследовании.

Респондент	Возраст	Счастье	Количество партнеров	Вероисповедание
1	30	[illegible]	[illegible]	[illegible]
2	32	1	1	Протестант
3	31	1	1	Католик
4	34	2	5	[illegible]
5	34	2	3	Протестант
6	3[illegible]	[illegible]	0	[illegible]
7	[illegible]	[illegible]	4	[illegible]
8	42	[illegible]	3	[illegible]
9	48	[illegible]	[illegible]	[illegible]
10	[illegible]	2	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	4[illegible]	2	[illegible]	[illegible]
[illegible]	29	[illegible]	8	[illegible]
[illegible]	28	[illegible]	[illegible]	[illegible]
15	47	2	[illegible]	[illegible]
16	69	[illegible]	2	[illegible]

Респондент	Возраст	Счастье	Количество партнеров	Вероисповедание
17	44	1	4	Другое
15	21	3	1	Протестант
19	33	2	1	Атеист
[illegible]	[illegible]	[illegible]	2	Протестант
[illegible]	[illegible]	[illegible]	0	Католик
22	51	1	1	Католик
[illegible]	[illegible]	2	[illegible]	Атеист
[illegible]	[illegible]	[illegible]	0	Протестант
25	47	2	3	Протестант
26	88	3	0	Католик
[illegible]	[illegible]	1	[illegible]	Протестант
28	24	1	1	Атеист
29	24	2	3	Атеист
[illegible]	[illegible]	[illegible]	[illegible]	Протестант

3.7. Найдите подходящий показатель центра распределения для каждой переменной, помещенной в таблицу 2.4. Опишите каждый показатель, как вы бы это сделали в официальном отчете об исследовании.

3.8. SOC

Для задачи 2.5 подберите наиболее подходящий показатель центра распределения для каждой переменной.

3.9. SOC

Власти рассматривают вопрос о полном запрете студенческих автомобилей. По данному вопросу вы провели опрос студентов и соседей, которые живут около кампуса и подсчитали баллы респондентов. В соответствии с использованной шкалой, высокое значение указывает на сильное несогласие с предложенным запретом. Ниже представлены значения обеих групп. Рассчитайте подходящий показатель центра распределения и в одном [illegible] [illegible]

Студенты		Соседи	
[illegible]	[illegible]	[illegible]	7
[illegible]	9	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	6
0	[illegible]	[illegible]	3
[illegible]	9	[illegible]	3
[illegible]	[illegible]	[illegible]	0
[illegible]	7	[illegible]	[illegible]
5	[illegible]	[illegible]	[illegible]
[illegible]	2	[illegible]	9
[illegible]	[illegible]	[illegible]	[illegible]

3.10 SW

Как глава агентства социального обеспечения вы полагаете, что ваш штат, состоящий из 20 социальных работников, слишком перегружен по сравнению с тем, что было 10 лет назад. Ниже приведено количество обслуживаемых каждым работником за интересующие два года соответственно. Увеличилось ли [illegible] обслуживаемых? Какой показатель центра распределения лучше всего подходит для ответа на этот вопрос? Почему?

1990		2000	
52	55	42	82
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	52	65	50
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	50	64	65
60	65	69	60
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
50	40	60	60

3.11 SOC

Рассчитайте медиану и среднее арифметическое для каждой из трех переменных, представленных в задаче 3.7. Есть ли переменные с положительной или отрицательной асимметрией? Как вы это определили?

3.12. SW

Для баллов по тесту, представленных в задаче 2.6, рассчитайте медиану и среднее арифметическое для предварительного и заключительного тестов. Интерпретируйте полученные показатели.

3.13 SOC

Выборка из 25 первокурсников в ведущем университете были участниками опроса, в котором измерялась их степень расовой нетерпимости (чем выше значение, тем выше нетерпимость).

а. Рассчитайте значения медианы и среднего арифметического для приведенных данных.

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

б. Те же студенты прошли тот же опрос в последний год обучения. Рассчитайте медиану и среднее арифметическое для второго набора значений и сравните их с предыдущими значениями. Что произошло?

[illegible]

3.14 [illegible] ... все рабочих в каждом городе [illegible] пользуются общественным транспортом [illegible] работы

Город	Рабочие, использующие общественный транспорт, %	Город	Рабочие, использующие общественный транспорт, %
[illegible]	[illegible]	[illegible]	2
		[illegible]	[illegible]
	[illegible]	[illegible]	2
[illegible]		[illegible]	[illegible]
		[illegible]	[illegible]
	[illegible]	[illegible]	9
Округ Колумбия	33	Сиэтл	[illegible]
		[illegible]	[illegible]
Сан-Франциско	36	Денвер	6
[illegible]	38	Канзас-Сити	
Кливленд	22	Миннеаполис	[illegible]

[illegible]

а. [illegible]

б. Сравните среднее и медиану. Какой из показателей [illegible]

в. Если вы удалите Нью-Йорк из распределения и пересчитаете показатели [illegible] ной)? Почему?

г. [illegible] как вы бы это сделали в оф[illegible] циальном отчете об исследовании.

[illegible] ...вкой, поскольку считают, что им слишком мало платят. Владельцы команд [illegible]

[illegible]

SPSS for Windows

Использование *SPSS for Windows* для вычисления центра распределения и процентилии

Запустите *SPSS for Windows*, нажав на иконку SPSS на экране монитора. Загрузите фа[illegible] и когда вы [illegible] к работе) на кнопке "закрывающей" экран, вы можете начать работать.

Демонстрационный пример с использованием SPSS 3.1. Определение показателей центра распределения

Единственной подпрограммой в SPSS, позволяющей получить все три наиболее часто используемых показателя центра распределения (моду, медиану и среднее арифметическое), является **Frequencies** (Частоты). Мы использовали данную подпрограмму, чтобы получить распределение частот в иллюстрациях 2.1 и 2.2 и в приложении Б. Здесь мы будем использовать **Frequencies** для расчета показателей центра распределения для трех переменных: *relig* (религиозные конфессии), *attend* (частота посещения церковных служб) и *age* (возраст).

Эти три переменные измерены по различным шкалам, и мы можем рассчитать лишь подходящий показатель центра распределения для каждой переменной. Таким образом, мы можем взять моду для *relig* (номинальная), медиану для *attend* (порядковая) и среднее арифметическое для *age* (интервальная). Хотя это и будет разумно, но в действительности гораздо удобнее получить все три показателя центра распределения для каждой переменной и проигнорировать неподходящий результат.

Чтобы получить моды, медианы и средние арифметические, начните, щелкнув **Analyze** (Анализ) в строке меню, и затем нажмите **Descriptive Statistics** (Описательная статистика) и **Frequencies**. В диалоговом окне **Frequencies** найдите названия переменных в списке слева и нажмите на кнопку со стрелкой посредине экрана, чтобы поместить названия (*relig*, *attend* и *age*) в окно **Variables** справа.

Чтобы рассчитать отдельные показатели, нажмите кнопку **Statistics** (Статистики) внизу диалогового окна **Frequencies**, и откроется диалоговое окно **Frequencies: Statistics** (Частоты: Статистики). Найдите прямоугольник **Central Tendency** (Центр распределения) справа и нажмите **Mean** (Среднее арифметическое), **Median** (Медиана) и **Mode** (Мода). Нажмите **Continue** (Продолжить) и вы вернетесь в диалоговое окно **Frequencies**, где вы возможно захотите нажать на блок **Display Frequency Tables** (Вывести таблицы частот). Когда этот блок *не* выбран, SPSS *не* выдаст таблицы частот, и появятся лишь показатели, которые мы указали (мода, медиана и среднее арифметическое), в окне **Output** (Вывод). Нажмите **OK** и SPSS выдаст следующее.

Показатели

		Возраст респондентов	Как часто респондент посещает церковные службы	Вероисповедание респондентов
Количество	Действительные	365	[illegible]	[illegible]
	Отсутствующие	2	[illegible]	[illegible]
Среднее арифметическое		44.94	3	3
Медиана		4[illegible].00	[illegible]	
Мода		34	0	

Обращая внимание лишь на наиболее подходящие показатели для каждой пе- [illegible] будет равна 1 (см. нижнюю строку [illegible] relig (Религиозные [illegible]). [illegible] это узнать либо обратившись к [illegible] Ж, [illegible]. Возс[illegible] Utilities [illegible] нажмите Variables [illegible] и найдите relig в списке переменных [illegible]. В любом случае, вы узнаете, что значение 1 означает *протестант*. Это [illegible] вероисповедания в выборке и, таким образом, ее модой.

[illegible] (как часто респондент посещает религиозные службы) имеет значение 3.00. Опять же, используйте либо приложение Ж, либо онлай[illegible] *несколько раз в год*. Это означает, что срединный случай в данном распределении [illegible] 365 наблюдений будет иметь значение 3 (или что срединное наблюдение находиться в интервале *несколько раз в год*).

Результат для *age* показывает, что респондентам в среднем было 44.94 года. Поскольку возраст является интервальной переменной, которая была измерена в [illegible] которое не нуждается в расшифровке, и нам нет нужды обращаться к кодировочной таблице для его интерпретации.

Обратите внимание, что SPSS не колеблется, рассчитывая средние [illegible] двух неинтервальных переменных. Программа не может отличить [illegible] кой (как для значений relig) и собственно число — для SPSS [illegible] решать, являются ли они верными с точки зрения статистики. Если вы затребуете бессмысленный показатель [illegible] SPSS выполнит вашу инструкцию без колебаний. Слепое желание SPSS просто [illegible] ваться командам упрощает для вас — пользователя запрос совершенно бес[illegible] ных показателей. [illegible] чение, они лишь перемалывают [illegible]

[illegible] Это не всегда будет [illegible] SPSS [illegible]

[illegible] страцию обратите внимание, что среднее [illegible]. Вернувшись к разделу 3.6, мы увидим, что [illegible] положительную асимметрию или несколько [illegible] мающих пожилым людям. Проверьте это, по[illegible] нажмите Graphs (Графики), затем нажмите Line [illegible] Simple Line [illegible] Define (Определить). Появится диалоговое окно Define [illegible] Category Axis [illegible] Options [illegible] нижнем левом углу и убеди-

[illegible] Line by groups defined by [illegible] нажмите Continue, а затем нажмите OK, и будет получен линейчатый график для *age*.

Обратите внимание на то, как график достигает максимума в возрастных группах [illegible] и как он снижается в группах 70 и 80. Хотя этот график не так плавен [illegible] нам, что *age* асимметричен в положительном [illegible] значений.

Демонстрационный пример с использованием SPSS 3.2. Команда Descriptives

Команда Descriptives ([illegible]) в *SPSS for Windows* разработана таким образом, чтобы предоставлять обобщающие показатели для непрерывных интервальных переменных. По умолчанию (т.е. если вы не задали иного), Descriptives рассчитает среднюю арифметическую, минимальное и максимальное значения (т.е. [illegible] значения) и стандартное отклонение (которое будет рассмотрено в главе 4). В отличие от Frequencies, эта подпрограмма не выдаст распределение частот.

[illegible] Descriptives [illegible] программу для переменных *age*, *educ* (Наибольшее количество лет обучения) и *tvhours* (Количество часов в день, проведенных у телевизора). Нажмите Analyze, Descriptive Statistics и Descriptives. Откроется диалоговое окно Descriptives. Это диалоговое окно выглядит точно также, как диалоговое окно Frequencies, и работает таким же образом. Найдите названия переменных в списке слева и, как только они были выделены, нажмите на кнопку со стрелкой посредине экрана, для того чтобы перенести их в окно Variables справа. Нажмите OK и вы получите следующий результат.

Показатели описательной статистики

	N	Минимум	Максимум	Среднее	Станд. отклон.
[illegible]та	1335	8	89	44.94	17.08
[illegible]	[illegible]	0	20	13.37	2.96
[illegible] телевизора	134	0	21	2.86	2.20
[illegible]	6				

В среднем в выборке представлены люди, которым 44.94 года (это повторяет результат Frequencies в иллюстрации 3[illegible]), которые проучились 13 лет и которые почти [illegible] часа каждый день проводят у телевизора.

Демонстрационный пример с использованием SPSS 3.3. Расчет процентилей, децилей и квартилей

[illegible] программа Frequencies может [illegible] (см. раздел 3.4). Эти показатели [illegible] децилей и квартилей для любой переменной [illegible] GSS, подходящих под это описание, [illegible] кой из немногих переменных в файле данных GSS, подходящих под это описание, является *age*, и мы еще один раз воспользуемся этой переменной для иллюстрации

[illegible]

Analyze → Descriptive Statistics → Frequencies [illegible] Statistics [illegible] Frequencies [illegible] Frequencies: Statistics [illegible] Percentile Values [illegible] Quartiles (квартили) и Cut points for 10 equal groups (Точки [illegible]). Благодаря последнему будут получены децили (десять [illegible])

[illegible] Percentiles [illegible] Add [illegible] Frequencies [illegible] Display Frequency Tables [illegible] OK [illegible] SPSS [illegible]

Возраст респондентов

Количество	Значимое	1385
	Недостающее	2
[illegible]		[illegible]
[illegible]		41.00
Мода		[illegible]
Процентили	10	[illegible]
	20	[illegible]
	[illegible]	[illegible]
	25	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	60	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]

[illegible]

Упражнения

3.1. Используйте **Frequencies** (SPSS, или **Univariate** (MicroCase) для того, чтобы получить все три показателя центра распределения для *marital*, *income98*, *race*, *region* [illegible] *prestg80* и для других пяти переменных по вашему выбору. Выберите наиболее подходящий показатель для каждой из этих переменных и напишите одно-два предложения, обобщая полученный результат

3.2. Используйте **Descriptives** (SPSS) или **Univariate** (MicroCase) для того, чтобы рассчитать средние арифметические для *partnrs5* и *sexfreq*. Обратите внимание, что эти переменные измерены по порядковой шкале и поэтому вычисления среднего арифметического не совсем обоснованны. Такое нарушение строгих правил уровня измерения обычно в социологических исследованиях. Напишите одно-два предложения, обобщая среднее арифметическое каждой переменной

3.3. (Только SPSS) [illegible] рассмотрите команды, которые [illegible] в упражнении 3.1 для того, чтобы рассчитать децили для каждой из перечисленных переменных

3.4. (Только MicroCase) Используйте опцию подкласса в диалоговом окне **Univariate** для сравнения мужчин и женщин по престижу занимаемого положения (*prestg80*) и посещению церкви (*attend*). Опишите полученные результаты в нескольких предложениях.

4 Меры разброса

Цели изучения Изучив данную главу, вы сможете

1. Объяснить назначение мер разброса и суть информации, которую они выражают.
2. Вычислить и объяснить смысл таких мер разброса: коэффициент изменчивости категорий (IQV), размах (R), интерквартильная широта (Q), стандартное отклонение (s), дисперсия (s^2).
3. Выбирать уместную меру разброса, правильно ее рассчиты[illegible]
4. Перечислить и объяснить математические свойства стандар[illegible]

4.1. Введение

[illegible] ...ние, имея выборку, следует использовать пару итоговых ста[illegible] **меру центра распределения** и **меру разброса**. Если среднее, медиана и мода [illegible] значения для представления часто встречающихся и/или центральных значе[illegible] распределения, то меры разброса дают нам информацию о степени неоднород[illegible] изменчивости внутри распределения значений.

[illegible] понять, почему понятие разброса так важно, рассмотрим [illegible] чиновник, отвечающий за общественную безопасность го[illegible] сравнить две службы скорой помощи, заключившие контракт [illegible] неотложной медицинской помощи. Как часть своего [illegible] данные о времени отклика на звонок о помощи для [illegible] отклика — это интервал времени от полу[illegible]

[illegible] времены отклика для первой и второй службы соответственно.

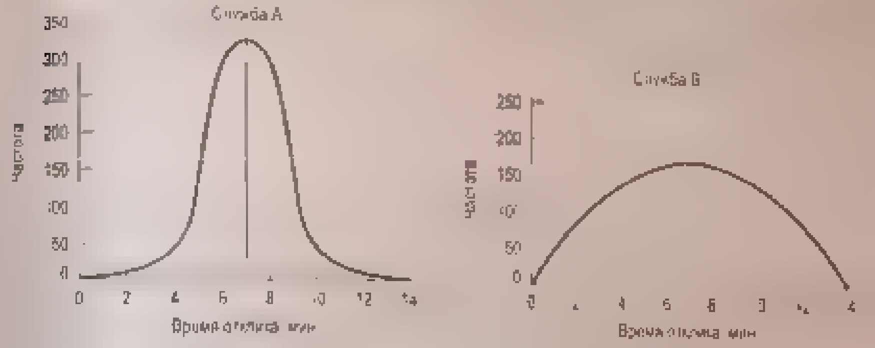


Рис. 4.1. *Распределения времени отклика для первой и второй службы скорой помощи*

Сравним форму двух кривых, изображенных на рис. 4.1. Заметим, что полигон частот для службы В намного более плоский, чем для службы А. Для службы В отдельные значения сильнее разбросаны по всему диапазону от минимального до максимального значения, в то время как для службы А отдельные значения собраны в кучу, сгруппированы вокруг среднего. Мы можем сделать вывод о том, что служба В подвержена намного большей изменчивости по времени отклика. Следовательно, несмотря на то, что оба распределения имеют практически одинаковые средние, распределение значений времени отклика для службы В является более неоднородным. Если бы мы не провели анализ разброса, то важное различие в эффективности работы двух служб скорой помощи осталось бы незамеченным.

Из рис. 4.1 вы уже составили себе общее понятие о том, что такое разброс, но [illegible] ве мы познакомим вас с некоторыми из наиболее распространенных в математической статистике мер разброса. Каждая мера разброса является точно определенным [illegible] циента изменчивости категорий, затем довольно кратко рассмотрим еще две меры разброса — размах и интерквартильную широту, а основное внимание уделим стандартному отклонению и дисперсии.

4.2. Коэффициент изменчивости категорий (IQV)

Коэффициент изменчивости категорий (IQV) [illegible]

[illegible] IQV [illegible]

Таблица 4.1. Расовый состав для трех районов проживания

Район A		Район B		Район C	
Расовая группа	Частота	Расовая группа	Частота	Расовая группа	Частота
[illegible]	[illegible]	Белые	50	Белые	[illegible]
[illegible]	[illegible]	Чернокожие	20	Чернокожие	30
[illegible]	[illegible]	Другие	10	Другие	[illegible]
	[illegible]		N = 90		N = 90

[illegible] суть рассматриваемой нами статистики, рассмотрим [illegible] трех районов. Так как все жители района A белые, то [illegible] Район C — самый неоднородный из трех районов. Давайте посмотрим, [illegible] IQV. Расчетная формула для IQV [illegible]

Формула 4.1 [illegible]

где [illegible]
N — количество наблюдений (объем выборки);
$\sum f^2$ — сумма квадратов частот.

[illegible] IQV по формуле 4.1, необходимо сначала найти сумму квадратов [illegible] в таблицу [illegible]

[illegible] IQV [illegible]

Таблица 4.2. Расчет сумм квадратов частот

Район A			Район B			Район C		
Расовая группа	Частота	Квадрат частоты	Расовая группа	Частота	Квадрат частоты	Расовая группа	Частота	Квадрат частоты
Белые	[illegible]	[illegible]	Белые	[illegible]	[illegible]	Белые	[illegible]	[illegible]
Чернокожие	0	0	Чернокожие	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Другие	[illegible]	0	Другие	[illegible]	[illegible]	Другие	[illegible]	[illegible]
	[illegible]	[illegible]		[illegible]	[illegible]		[illegible]	[illegible]

Так как [illegible] IQV [illegible] районов можно найти, подставив в формулу для района A другое значение. Так, для района B

$$IQV = \frac{[illegible]}{8100[illegible]}$$

$$IQV = \frac{[illegible]}{16\,200}$$

$$IQV = [illegible]$$

аналогично для района C

$$IQV = \frac{[illegible]}{[illegible]}$$

$$IQV = \frac{[illegible]}{[illegible]}$$

$$IQV = 1{,}00.$$

Итак, используя IQV, мы подтвердили наши рассуждения об изменчивости для [illegible] показывает максимально возможное для наших данных количество изменчивости (IQV = 1,00). *Чтобы попрактиковаться в вычислении и интерпретации коэффициента изменчивости категорий IQV, решите задачи 4.1 и 4.2.*

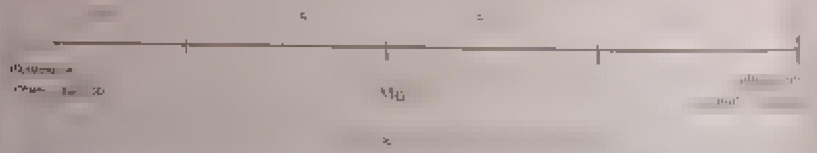
4.3. Размах (R) и интерквартильная широта (Q)

По определению **размах** (*R*, от "range") — это [illegible] наименьшим наблюдаемым значением для распределения [illegible] ковой или интервальной переменной. Вычисление [illegible] читания наименьшего наблюдаемого значения [illegible]

[illegible] часто вводит нас в заб[illegible]

[illegible] чрезмерно больших или малых значений, то размах [illegible] R [illegible]

Интерквартильная широта (Q, от "interquartile range") — это тоже рас[illegible] между верхним и нижним квартилями. [illegible] для размаха R, т.к. рассматривает только средние 50% наблюдений выборки. Чтобы найти Q, нужно отсортировать наблюдения по возрастанию, т.е. [illegible] вариационный ряд. Затем найти квартили — величины, делящие выборку [illegible] группы, содержащие (по возможности) одинаковые количества наблюдений (А[illegible] логично тому, как мы разбивали выборку на две группы, чтобы найти медиану). Нижняя или первая квартиль (Q_1) — это такая точка, ниже которой в упорядоченном множестве данных находится 25% данных и, соответственно, выше которого [illegible] 75% данных. Вторая квартиль (Q_2) — это такая точка, по обе стороны [illegible] торой располагается одинаковое количество элементов выборки: очевидно, Q_2 является медианой. Верхняя или третья квартиль (Q_3) — это такая точка, ниже которой в упорядоченном множестве данных находится 75% данных и, соответственно, выше которого находится 25% данных. Пусть отрезок LH представляет собой распределение значений; нанесем на него значения квартилей.



[illegible] какую-либо информацию о природе тех [illegible] отличных от первого и третьего квартиля.

Таблица 4.3. Расходы на государственное образование в 1998 г. на душу населения

Штат	Расходы на душу населения
Нью-Джерси	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

Штат	Расходы на душу населения
[illegible]	[illegible]
Иллинойс	1147
Нью-Гэмпшир	1142
Пенсильвания	1140
Вирджиния	1132
Айдахо	1122
Небраска	1117
Флорида	[illegible]
Аризона	1064
Калифорния	1061
Северная Каролина	1012
Алабама	1011
Луизиана	95[illegible]
Миссисипи	939

Источник: U.S. Census Bureau, [illegible] Statistical Abstract [illegible] edition). Washington, D.C., 2000, p. 182

4.4. Расчет размаха и интерквартильной широты[1]

В таблице 4.3 представлены данные о расходы на государственное школьное образование на душу населения для 20 штатов. Каков размах и интерквартильная широта для этой выборки?

Отметим, что значения интересующей нас переменной *расход на душу населения* уже отсортированы по убыванию. Поэтому мы можем сразу же рассчитать размах, а при расчете интерквартильной широты нам не придется сортировать наблюдения. Для 20 штатов, перечисленных в таблице 4.3, Нью-Джерси тратит наибольшее количество средств на государственное образование на душу населения ($1571), а Миссисипи — наименьшее ($939). Следовательно, размах равняется 1571−939=632 (R=$632).

Чтобы найти Q, мы должны сначала найти первый и третий квартиль (Q_1 и Q_3). Аналогично тому, как мы поступали при нахождении медианы, привяжем нужные нам точки к конкретным наблюдениям выборки. Так, найдем Q_1 путем умножения N на (0,25): (20)×(0,25)=5, поэтому Q_1 равняется значению наблюдения номер 5, считая от наименьшего наблюдения. Наблюдение номер пять — это [illegible] [illegible] нахождения третьего квартиля отыщем соответствующее наблюдение, [illegible] (0,75) = 15 [illegible] Наблюдение № 15 [illegible] дохода на душу населения 1206. Поэтому Q_3=1206. Итак,

$$Q = Q_3 - Q_1$$

[illegible]

[1] Этот раздел не обязателен.

[illegible] Q_1 [illegible] не будет таким очевидным. Как [illegible] т.е. первая квартиль будет равняться значе[illegible] [illegible] 25. А для Q_3 имеем (57)×(0,75), т.е. третья квартиль [illegible] с наблюдением под номером 117,75. Наблю[illegible] непозможны, поэтому возникает проблема. Самое про[illegible]стое решение этой проблемы — взять значение наблюдения с номером, наиболее близким к рассчитанному нами номеру. Тогда Q_1 будет равняться значению 39-го наблюдения, а Q_3 — значению 118-го наблюдения.

Более точное решение состоит в том, чтобы учесть дробные части номеров наблюдений. Например, Q_1 можно определить как такое значение, которое больше значения 39-го наблюдения на величину, равную 1/4 части расстояния между 39-м и 40-м наблюдениями. Аналогично Q_3 можно определить как такое значение, которое больше значения 117-го наблюдения на величину, равную 3/4 части расстояния между 117-м и 118-м наблюдениями. Описанная процедура похожа на процедуру расчета медианы, которая, между прочим, является второй квартилью. Если объем выборки нечетен и равен $2n+1$, то медиана равна элементу вариационного ряда с номером $n+1$. Если объем выборки четен и равен $2n$, то медиана лежит между элементами вариационного ряда с номерами n и $n+1$, как правило, в таких случаях медианой считают среднее арифметическое этих двух значений. В большинстве задач различия между двумя значениями Q — полученными, соответственно, первым и вторым, более точным методом, — крайне малы. *Чтобы попрактиковаться в вычислении и интерпретации интерквартильной ширины Q, решите задачу 4.5. Вообще, рассчитать Q возможно для любой задачи этой главы, кроме 4.1 и 4.2.*

4.5. Стандартное отклонение

Основным недостатком R и Q является то, что [illegible] не используются все наблюдения данной выборки — т.е. эти статистики не отражают всю доступную информацию. [illegible]

1. Использовать все значения данного распределения.
2. Описывать среднее [illegible] отклонение данных [illegible]
3. [illegible]

[illegible]

[illegible] от [illegible] наблюдений от среднего?

Мож[illegible] $\sum$ [illegible] ратся [illegible] 20, [illegible], 40, 50. Просуммировав отклонения наблюдений от среднего, получим:

Значения	Отклонения	
10	(10 − 30) =	−20
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
40	(40 − 30)	[illegible]
[illegible]	[illegible]	[illegible]
$\sum$ = 150	$\sum$ [illegible]	

$\bar{X} = 150/5 = 30$

И все же сумма отклонений является логической основой для статистики, которая бы измеряла количество изменчивости для данного набора наблюдений, и в математической статистике известны [illegible] решения [illegible], что сумма положительных отклонений всегда равна сумме отрицательных отклонений. Итак, оба решения устраняют знак минус. Первое [illegible] в использовании модуля, т.е. игнорировании знака каждого отклонения при вычислении суммы. В нашем примере при игнорировании знака мы получим сумму, равную (20+10+0+10+20)=60.

Второе решение заключается в использовании квадрата. Благодаря свойству операции возведения в квадрат все слагаемые становятся неотрицательными. В нашем примере сумма квадратов отклонений составит (400+100+0+100+400)=1000.

Как видим, у нас есть две стратегии для разработки "хорошей" меры разброса. Давайте введем еще одно усовершенствование. При использовании абсолютных [illegible] или квадратов отклонений сумма будет возрастать с ростом объема выборки [illegible] наблюдений в выборке, тем больше будет значение нашей меры разброса). Это затруднит сравнительный анализ изменчивости для выборок различного объема. Чтобы решить такую проблему, давайте разделим сумму квадратов от[illegible]

[illegible]

[illegible] средним отклонением (AD) [illegible]

Формула 4.2 $AD = \frac{\sum [illegible]}{N}$

[illegible] что мы должны складывать отклонения без учета знака

Для рассматриваемого примера AD составит

$$AD = [illegible]$$

$$AD = \frac{60}{5} = 12$$

Это среднее отклонение говорит о среднем [illegible]

[illegible] отвечает всем трем критериям "хорошей" меры разброса, выдвинутым нами ранее. Но из-за трудностей алгебраических операций с абсолютными значениями и, следовательно, трудностей анализа ее выборочных свойств эта мера практически не используется, в том числе и в социологии.

Мера разброса, устраняющая знак минус путем возведения в квадрат каждого отклонения, и соответствующая статистика называются дисперсией. Дисперсия выборки обозначается s^2, а дисперсия генеральной совокупности — σ^2 (строчная греческая буква "сигма" в квадрате). Дисперсия используется в основном в статистических выводах, а также является центральным понятием при конструировании некоторых мер связи. Чтобы описать разброс распределения, используется другая статистика, тесно связанная с дисперсией, — стандартное отклонение. Стандартное отклонение выборки обычно обозначается s, а стандартное отклонение генеральной совокупности — σ. И именно стандартному отклонению будет посвящена оставшаяся часть этой главы. Формулы для дисперсии выборки и стандартного отклонения выборки таковы

Формула 4.3 $[illegible]$

Формула 4.4 $[illegible]$

Строго говоря, формулы 4.3 и 4.4 предназначены для расчета дисперсии и стандартного отклонения генеральной совокупности. А при работе не с генеральными [illegible] выборками для получения несмещенной оценки [illegible]

программное обеспечение, которым [illegible] что дает вам результат [illegible] полученного по формулам 4.3 и 4.4. Некоторые [illegible] менатель ($N-1$) или N. Если вы выберите N, то тогда получите в точности те [illegible] значения, что и в примерах этой главы.

В любом случае, давайте рассчитаем стандартное отклонение для нашего примера.

$$[illegible]$$

[illegible]

4.14

Таблица 4.4. Расчет стандартного отклонения по формуле-определению [illegible] формулы 4.3

Балл студента X	[illegible]	[illegible]
[illegible]		
[illegible]		
70		
[illegible]		
[illegible]		
[illegible]	[illegible]	
[illegible]		
[illegible]	[illegible]	
[illegible]		
$\sum$ [illegible]	$\sum$ [illegible]	$\sum$ [illegible]

[illegible]

[illegible] дисперсия равна квадрату стандартного отклонения [illegible] этого примера дисперсия [illegible]

4.6. Удобная формула для расчета стандартного отклонения

Формула 4.4 — это формула-определение стандартного отклонения [illegible] расчета. Для ускорения и упрощения вычислений предложено большое количество формул на основе формулы 4.4. Наиболее распространенная [illegible]

Формула 4.5 $[illegible]$

[illegible]

[illegible]

совершенно различных вычислений. [illegible] путем возведения каждого значения в квад[illegible]

[illegible]

возведения в квадрат. См. раздел "Basic Math Review" (обзор основ математики) на веб-сайте данной книги, где различие между этими двумя обозначениями объясне-но более подробно.

Для иллюстрации давайте проведем расчет по формуле-определению 4.4 и по удобной для расчета формуле 4.5. В таблице 4.4 представлены баллы, набранные 10 студентами при проверке их знаний по статистике. Подставив в формулу 4.4 циф-ры, рассчитанные в таблице 4.4, имеем:

[illegible]

Таблица 4.5. Расчет стандартного отклонения по формуле, удобной для расчета, формуле 4.5

Балл студента (X_i)	X^2
[illegible]	3249
[illegible]	3600
[illegible]	4225
[illegible]	[illegible]
[illegible]	6084
[illegible]	6400
[illegible]	6724
[illegible]	7225
[illegible]	8100
[illegible]	8649
	$\sum X_i^2 = 59\,656$

[illegible] вычитания среднего из каждого наблю-[illegible] вычисления. Подставив в формулу 4.5 [illegible]

[illegible]

Заметим, что при расчете по формуле 4.5 у нас получилось в точности то же зна-[illegible] формулы 4.4, но позволяет значительно упростить вычисления. Заметим также, что [illegible] ре дисперсии составит $(11{,}82)^2 = 139{,}71$. *Чтобы попрактиковаться в вычислении и интерпретации s и s^2, решите любые задачи в конце этой главы, кроме задач 4.1 и 4.2. Сна-чала рекомендуем задачи с меньшими объемами выборок — например, 4.3 и 4.5, а затем* [illegible]

Практический пример 4.1

[illegible] этом подсчитали количество актов агрессии, совершенных каждым ребенком [illegible] клонения воспользуемся формулой 4.5.

Количество актов агрессии

X_i	X_i^2
[illegible]	1
0	0
3	9
5	25
2	4
7	49
0	100
0	0
3	64
2	4
0	0
$\sum X_i =$ [illegible]	$\sum X_i^2 =$ [illegible]

[illegible] 4.5 получим

[illegible]

[illegible]тандартное отклонение для наших данных равняется 3,44

Практический пример 4.2

Вы только что выиграли в лотерею [illegible] Даже после уплаты налогов у вас осталось достаточно денег, что[illegible] законное и совершенное [illegible] такие [illegible] рассуждения, как не[illegible] нести работать. Одним из ваших первых решений будет отыскать самое [illegible] место для проживания, во всем мире. А т.к. вы немного эксцентричны [illegible] что единственный критерий самого лучшего места — это кл[illegible] [illegible] в точности 74°F. Какой из этих трех городов вы [illegible] в качестве по[illegible] места жительства [illegible]

[illegible] дневной температуры для трех городов: 0,7 и 3° для города А; 10 [illegible] 25,6 и 102° для города С. Соответственно сможете ли вы теперь [illegible] место постоянного жительства? Какой город вы выберите и почему?

4.7 Интерпретация стандартного отклонения

[illegible] что вам до сих пор непонятно, зачем мы рассчитываем стан[illegible] отклонение. Вы [illegible] "Я [illegible] посчитал стандартное отклонение, затра[illegible] [illegible] важный из них основан на применении [illegible]

Втор[illegible] том, чтобы считать стандартное отклонение [illegible] [illegible] том изменчивости [illegible] [illegible] рого отсутствует и[illegible] наблюдения равна между собой. В этом случае стандартное отклонение будет равно нулю. Следовательно, 0 [illegible] наименьшее возможное значение для стандартного отклонения (хотя верхнего пре[illegible])

[illegible] интерпретации стандартного отклонения заключается в сравнении одного распределения с другим. Пусть у нас есть следующая информация о распределении дохода на одно домовладение для двух городов соответственно

Город А	Город B
$\bar{x}$ = 31,255	[illegible]
[illegible]	[illegible]

Эти два города не отличаются по центру распределения, но изменчивость дохода от домовладения к домовладению намного больше для города B. Это может быть вызвано тем, например, что в городе B больше семей из кат[illegible] а также из категории с очень низким доходом по сравнению с тем, что семьи в городе А сильнее сгруппированы около среднего дохода. Итак, стандартное отклонение [illegible] для сравнения нескольких распределений

4.8 Интерпретация статистик: использование мер центра распределения и мер разброса для анализа этнической неоднородности населения США

В тот момент, когда США вступили в XXI век, неоднородность уже перестала быть просто вопросом политической корректности: она стала социальной и [illegible] кой реальностью. Используя данные Бюро переписи населения США, этот раздел, посвященный интерпретации статистик, описывает расовый и этнический состав 50-ти штатов и округа Колумбия в 2025 году.

Тенденции в процентной доле расово-этнических групп, для США в целом

Изменения в процентном соотношении пяти расово-этнических групп [illegible] времени, свидетельствует о росте [illegible] ния США (см. табл. 4.6). В этой таблице представлены итоговые статистики [illegible]

[illegible]

[illegible] к числовой шкале, а именно — к ее разновидности *шкала отноше-* [illegible] этой шкалы есть естественное *начало* (значение этой пе- [illegible] этнической группы) и соблюдается равенство интервалов.

[illegible] меры разброса

Ранее мы говорили, что исходные выборки и таблицы частот могут не приводиться в сообщениях исследователей в профессиональной литературе. Аналогично могут не приводиться меры центра распределения и меры разброса. Ведь в типовом исследовательском проекте задействовано множество переменных, а объем статей в журналах и книгах место ограничен. Поэтому исследователю просто не хватает места, чтобы полностью описать каждую переменную. Более того, большая часть сообщений об исследованиях уделяет внимание связям между переменными, а не распределению значений одной переменной. В этом случае описательные статистики одной переменной не относятся к главной цели сообщения.

Конечно, сказанное не означает, что описательные статистики одной переменной не важны для исследователя. Я совсем не имел в виду, что исследователям не нужно рассчитывать и интерпретировать эти статистики. В каждом проекте обязательно рассчи[illegible] и интерпретируются меры центра распределения и меры разброса, причем [illegible] переменной. Однако вероятность того, что эти статистики сообщат в окон[illegible] отчете об исследовании меньше, чем для данных о применении более слож[illegible] ческих методов, которые мы опишем позднее в [illegible] другой [illegible] в первой [illegible] во второй [illegible] либо сводки по всем нужным переменным — [illegible] таблицы. Средние и стандартные отклонения большого количе[illegible], например, представить в таком формате:

Переменная	$\overline{X}$	s	N
[illegible]	[illegible]	[illegible]	[illegible]

Такие [illegible] неполнотой информации по некоторым наблюдениям.

Описательные статистики в профессиональной литературе

В этом разделе мы приведем две краткие выдержки из двух социологических исследований. В этих статьях описательные статистики не являлись целью исследования, а были основой для статистического критерия, который мы рассмотрим в главе 9. В первой статье приводятся средние и стандартные отклонения, а во второй — только средние.

Выдержка 1. Различия в продуктивности ученых, вызванные полом

Профессор Дж. Скот Лонг (J. Scott Long) заметил, что типичная женщина-ученый публикует меньше статей и на ее работы меньше ссылаются, чем типичный мужчина-ученый. Эта тенденция наблюдается для различных дисциплин и не меняется с течением времени (p. 1297). Какими факторами вызвана низкая продуктивность женщин-ученых? Пытаясь это выяснить, профессор Лонг сформировал выборку объемом 556 мужчин и 603 женщины, получивших степень доктора биохимии в промежутке между 1950–1967 гг. Он собрал информацию о научных карьерах этих людей, их продуктивности как ученых и личной жизни. Некоторые из результатов представлены в таблице ниже. Внимательно изучите эту таблицу и попробуйте выделить тенденции из представленной информации.

Различия в значениях основных переменных, вызванные полом

	Пол	Среднее	Стандартное отклонение
[illegible]	Ж	[illegible]	[illegible]
	М	[illegible]	[illegible]
[illegible]	Ж	[illegible]	[illegible]
	М	[illegible]	[illegible]
[illegible]	Ж	[illegible]	[illegible]
	М	[illegible]	[illegible]
[illegible]	Ж	2,456	[illegible]
	М	[illegible]	[illegible]
[illegible]	Ж	0,26[illegible]	[illegible]
	М	0,3[illegible]	[illegible]
[illegible]	Ж	0,[illegible]	[illegible]
	М	[illegible]	[illegible]

Профессор Лонг использует средние и стандартные отклонения и в первом, и во вто[illegible] описательные статистики [illegible] статистического критерия, помогающего обнаружить важные и закономерные различия в значениях переменных, вызванные полом. [illegible]

[illegible]

[illegible] мужчин, [illegible] [illegible] для старше мужчин. Более высокий возраст [illegible]

[illegible]

[illegible] делает вывод: основной фактор, вызывающий различие в продуктивн[illegible] [illegible] и мужчин-ученых — это то, что женщины меньше работают со [illegible] [illegible] чаще мужчин прерывая свое обучение (на время беременн[illegible] [illegible] и ухода за ребенком). Профессор Лонг говорит: "Женщины с маленькими детьми [illegible] реже посещают своего научного руководителя, что вызывает снижение [illegible] женщины ученого во время работы над докторской диссертацией". [illegible] [illegible] хотите узнать больше? Ниже приведена полная библиографичес[illegible]

[illegible]

Выдержка 2. Адаптация в США иммигрантов из Африки: исследование и практика социальной работы

Профессор Хьюго Камия (Hugo Kamya) отмечает, что исследователи, рассматривая [illegible] эмигрантов, практически не рассматривают группы иммигрантов из Африки. В [illegible] этого социальным работникам и другим организациям сферы обществен[illegible] [illegible]

[illegible] Автора особенно заинтересовала роль религии в процессах адаптации в [illegible] влияющий на свою жизнь, второй — сильно влияющий). Ниже приведе[illegible] результаты.

Исследование влияния религии на следующие переменные, измеренные в баллах [illegible] способность бороться с стрессом, стресс, самооценка, душевное благополучие, [illegible] противодействия стрессу. Приведены средние по каждой переменной [illegible]

	[illegible] слабо [illegible]	[illegible] (N = 37)
[illegible]	[illegible]	41.7

Вот что пишет пр[illegible]

[illegible] [illegible] угроза культурным или религиозным верованиям [illegible] снижают возможность противодействия стрессу данной личности".

Вы хотите узнать больше? Ниже приведена библиографическая ссылка [illegible]

[illegible] Social Work, 42, 154 [illegible]

Давайте внимательно посмотрим на таблицу 4.6. Можете ли вы [illegible] выявить основные тенденции? Это может потребовать некоторых усилий, т.к. таблица "перегружена" большим количеством информации. По мере того, как вы читаете таблицу, вспомните определения статистик: среднее — это среднее арифметическое наблюдений, медиана — это значение наблюдения, по обе стороны от которого расположено равное количество наблюдений. Размах — это расстояние между наибольшим и наименьшим наблюдением, а стандартное отклонение, как и размах, измеряет количество разброса или изменчивости наблюдений. Не пытайтесь сразу "переварить" всю таблицу. Ваша задача — просмотреть ее и выявить общие тенденции.

Таблица 4.6. Итоговые статистики, характеризующие расово-этническую неоднородность, 1990–2025

Оценка для года	1990	1995	2000	2005	2015	2025
[illegible]						
[illegible]	84.76	84.13	83.47	82.82	81.55	80.33
[illegible]	88.24	87.56	87.09	86.31	84.93	83.61
[illegible]						
[illegible]	66.80	65.54	64.73	64.10	65.21	66.41
[illegible]						
[illegible]	10.76	[illegible]	11.30	11.56	12.10	12.63
[illegible]	7.18	7.12	7.37	7.30	8.95	8.74
[illegible]	2.11	11.01	[illegible]	1.83	11.34	12.17
[illegible]						
Медиана	1.63	1.62	1.64	1.66	1.76	1.86

Стандартное отклонение
Размах

[illegible] доля американцев [illegible] происхождение

[illegible]	1990	1995	2000	2005	2015	2025
[illegible]	62,59	63.06	62 78	62 91	63 42	65 02
[illegible]						
[illegible]	5,40	6.05	6,76	7,44	8 76	[illegible]
[illegible]	2,39	3,05	3,58	4 08	4 81	[illegible]
[illegible]	7.49	7 95	8 42	8,95	9 99	[illegible]
[illegible]	37 98	38 53	38 95	39 99	42 97	[illegible]
[illegible]	51.00	51.00	51.00	51 00	51.00	[illegible]

[illegible] давайте начнем с анализа мер центра распределения. С их помощью мы обнаруживаем две тенденции. Первая и, наверное, наиболее значимая тенден- [illegible] белого населения монотонно убывает с течением времени. Доля белых амери- [illegible] в населении США снизится в XXI веке. Во-вторых, для той же переменной [illegible] *ая доля белого населения* для всех шести лет среднее меньше медианы. Это [illegible] ает, что у данной переменной отрицательный коэффициент асимметрии (см. [illegible] 3 б), т.е. имеется несколько наблюдений с очень низкими значениями. Если [illegible] посмотреть на исходные выборки для штатов (которые мы здесь не приво- [illegible] что такая асимметрия является резул [illegible] округа Колумбия и Гавайи. В то время [illegible]

[illegible] переменной *процентная доля* [illegible] ное поведение отдельных значений (то, что [illegible] значения переменной, а некоторые — очень [illegible]

[illegible] ение можно считать способ- [illegible] значение, тем больше изменчи [illegible]

[illegible] ность стандартного отклонения переменной *процентная доля б* [illegible] ря о том, что количество [illegible] между штатами в 2025 г. будет оставаться примерно таким же, как и в 1990 г.

[illegible]

ся, превысив 10%.

[illegible] и среднее). Для всех четырех "не белых" групп среднее больше медианы, указывая на *положительный* коэффициент асимметрии соответствующих переменных. Отмеченная тенденция прямо противоположна тенденции для переменной *процентная доля белого населения*, и особенно ярко проявляется для переменных *процентная доля индейцев* и *процентная доля американцев азиатского происхождения*. Положительный коэффициент асимметрии означает, что в то время как большинство штатов имеют малый процент индейцев и жителей азиатского происхождения, все же есть несколько штатов с относительной высокой долей таких жителей, и эти штаты "вытягивают" среднее вверх, по сравнению с медианой.

Наконец, рассмотрим разброс. В целях экономии места мы не будем анализировать все группы, но давайте в качестве иллюстрации рассмотрим тенденции переменной *процентная доля американцев испанского происхождения*. Обе меры разброса — стандартное отклонение и размах — подтверждают наличие хорошо заметной тенденции к увеличению изменчивости. Так, стандартное отклонение по этой переменной возрастает от 7,49 в 1990 г. до 11,03 в 2025 г., аналогично размах возрастает от 37,98 до 46,29. Другими словами, все 50 штатов и округ Колумбия становятся более неодно- [illegible] Что вызвано? Некоторые штаты становятся "более испанскими" (т.е. в них с течением времени наблюдается рост доли американцев испанского происхождения), а то [illegible]

[illegible] пример [illegible] *значения* остаются [illegible]

[illegible] стандартного отклонения. Попробуйте [illegible] [illegible] всех групп [illegible] о информации, приведенной в табли- [illegible] США стоит перед лицом значительных перемен. В ближайшем будущем [illegible] [illegible] преобладание белого населения уменьшится. Другие группы (особенно группы аме [illegible] [illegible] [illegible] [illegible] снизится до 2025 г. Некоторые штаты оста[illegible] (95% или более белого населения), в других же штатах процент белого населения будет продолжать [illegible] ся в меньшинстве, — увеличиваться.

Тенденции в процентной доле расово-этнических групп по регионам

Более близкое знакомство с этим подразделом данной главы покажет вам, что тенденция к увеличению расово-этнической неоднородности наблюдается не во всех штатах. Что мы можем узнать путем объединения штатов в регионы и расчета средних по каждой расово-этнической группе для различных годов? В таблице 4.7 показаны средние для переменных *процентная доля данной расово-этнической группы в данном регионе* для трех лет: 1990, 2005, 2025 г.

Если вы будете часто иметь дело с описательными статистиками, то выработаете [illegible]

Таблица 4.7. Средние для переменных *процентная доля данной расово-этнической группы в данном регионе* [illegible]

Географический [illegible]	1990 [illegible]	1990 Стандартное [illegible]	2005 Среднее	2005 [illegible]	2025 Среднее	2025 [illegible]
Северо-Восток						
[illegible]						
американцев (индейцев)	0,3	0,1	0,3	0,1	0,4	0,2
[illegible]						
Средний Запад						
[illegible]						
Юг						
[illegible]	0,8	1,8	0,8	2,0		
[illegible]						
Объем выборки (количество штатов в регионе)	17,0		17,0			
Запад						
% белого населения	86,0	6,8	84,0	6,[illegible]		
[illegible]			7,6	2,3		
[illegible]	3,8					
[illegible]						

Применение коэффициента изменчивости категорий (IQV)

В таблице 4.7 [illegible] меры разброса [illegible] Но мы сможем сделать вывод о росте расово-этнической неоднородности [illegible] США и в том случае, если будем анализировать значения коэффициента [illegible] смотренных нами, которую можно рассчитать [illegible] человек [illegible] происхождения может быть отнесен к любой из расовых групп (белые, афро-американцы, индейцы, азиатского происхождения), то давайте рассмотрим переменную *расовый состав населения*, принадлежащую [illegible] и имеющую четыре категории. В таблице 4.8 представлены для 1990, 1995, 2000 г. наблюдаемые частоты различных расовых групп, полученные по реальным данным, для 2005, 2015 и 2025 г. — оценки частот, полученные путем прогнозирования. Вспомним, что чем выше IQV, тем больше изменчивость в данном распределении. Таблица 4.8 ясно показывает, что за период 1990–2025 г. в США будет происходить монотонный рост изменчивости переменной *расовый состав населения*.

Как пел Боб Дилан, *времена меняются*, и это высказывание наверняка применимо к расовому и этническому составу населения США. На основе приведенных здесь статистик мы делаем вывод: расовая и этническая неоднородность будет возрастать в каждом регионе, и к 2025 г. США будут иметь состав, сильно отличающийся от прежнего.

Таблица 4.8. Оценки расового состава населения США, 1990–2025 гг.

	Частоты (тыс. человек), для года					
	1990	1995	2000	2005	2015	2025
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Американцы азиатского происхождения	7,565	9,349	11,245	13 2[illegible]2	[illegible]	[illegible]
Итого (объем выборки)	249,464	262,744	274,621	285,9[illegible]	[illegible]	[illegible]
Коэффициент [illegible]	0,[illegible]4	0,39[illegible]	0,4095	0,42[illegible]	[illegible]	0,[illegible]

Замечание по поводу расовых и этнических категорий

Расовые категории, использованные нами в разделе 4.8, а также [illegible] расового состава США, не говоря уже о [illegible] категории не оставляют места американцам [illegible] расового и этнического

населения. Сам факт споров [illegible]

Заключение

1. Меры разброса — это итоговые статистические показатели, обобщающие информацию о неоднородности или изменчивости распределения значений [illegible] переменной. [illegible] центра распределения эти статистики [illegible] дают нам центральные точки распределения, то меры разброса сообщают, како[illegible] распределения.
2. Коэффициент изменчивости категорий (IQV) можно рассчитать для переменной, принадлежащей к любой шкале, но, например, для числовой переменной необходимо предварительно сгруппировать [illegible] частот классов. IQV — [illegible] наблюдаемого в данном распределении к максимально возможной в данном распределении изменчивости. IQV чаще всего вычисляется для переменных, принадлежащих к номинальной шкале (кстати, это единственная мера разброса, применимая к ним).
3. Размах (R) — это расстояние от наибол[illegible] Интерквартильная широта (Q) [illegible] (т.е. диапазон, в который попадают средние по величине 50% наблюдений выборки). Размах и интерквартильную широту можно ра[illegible] бых переменных, кроме принадлежащих к номинальной шкале.
4. Стандартное отклонение (s) — это самая важная мера разброса [illegible] роль во многих сложных статистических методах. Минимально [illegible] (если $s = 0$, то это указывает на отсутствие изменчивости в распределении), рост значения стандартного отклонения происходит по мере роста изменчивости распределения. Наиболее часто стандартное отклонение вычисляют для переменных [illegible] интервальн[illegible]
5. Дисперсия (s^2) в основном используется в статистических выводах и [illegible] конструировании некоторых мер связи.

Сводка формул

Коэффициент изменчивости категорий [illegible]

Среднее отклонение [illegible]

[illegible]

Стандартное отклонение (выборки) 4.4 $s=\sqrt{\frac{\sum(X_i-\bar{X})^2}{N}}$

[illegible] 4.5 $=\frac{1}{N}\sqrt{N\sum X_i^2-(\sum X_i)^2}$

Словарь

Дисперсия [illegible] на N. Эта мера разброса используется [illegible] в методах корреляционного и регрессионного анализа. Дисперсия выборки обозначается s^2, а дисперсия генеральной совокупности — σ^2.

Интерквартильная широта (**Q**, от interquartile range). Расстояние между третьей и первой квартилью.

Коэффициент изменчивости категорий (**IQV**, от index of qualitative variation). Мера разброса для таких переменных, которые представлены категориями и их на[illegible]

Меры разброса. Статистические показатели, характеризующие степень [illegible]ности или неоднородности в распределении значений интересующей [illegible]

Отклонение. Расстояния между наблюдениями выборки [illegible]

Разброс. Количество и [illegible]

Размах (**R**, от "range"). Результат вычитания наименьшего наблюдаемого [illegible]

Среднее отклонение (**AD**, от "average deviation"). Среднее арифметическое [illegible] ных отклонений наблюдений выборки от среднего.

Стандартное отклонение [illegible] рки от среднего), деленный на N. Это самая важная и полезная описательная мера разброса [illegible] отклонение выборки обычно [illegible]

Мультимедийные ресурсы

The Wadsworth Sociology Resource Center: Virtual Society (Виртуальное сообщество центра социологических ресурсов Вадсворта)

http://sociology.wadsworth.com

[illegible] шестое издание книги *Statistics: A Tool for Social* [illegible] (*Статистика как инструмент социологических исследований*), где вы найдете *Student Resources* [illegible] статистику. Начните со щелчка на [illegible], чтобы получить доступ к таким материалам

- [illegible]
- [illegible]
- [illegible]
- [illegible]
- Дополнительные [illegible]
- [illegible]
- [illegible] использованием пакетов статистической обработ[illegible]
- [illegible]
- [illegible]

Задачи

4.1. **SOC**

[illegible] данные о семейном положении жителей квартир четырех жи[illegible]. Рассчитайте коэффициент изменчивости категорий (IQV) для [illegible] комплекса A–D. Какой из этих четырех комплексов наиболее неодно[illegible] *за образец таблицу 4.2*

[illegible]

4.2. **SOC/CJ**

Рассчитайте коэффициент изменчивости категорий (IQV) для распределени[illegible]

4.3. Рассчитайте [illegible]

4.4. [illegible] 95 [illegible] 55, 45

4.5. **GER**

Рассчитайте размах, интерквартильную широту, стан[illegible]персию для данных задачи 2.8

4.14. [illegible]
Рассчитайте стандартное отклонение для первого и второго набора данных (со [illegible] и годам) задачи 3.10. Сравните эти стандартные отклонения, не забыв сравнить и средние. Что происходит и почему?

4.15. [illegible] первого и второго набора данных (со [illegible] стандартное отклонение [illegible] причиной, вызвавшей изменение стандартного отклонения? Что при этом про[illegible]

4.16. На математическом факультете колледжа Святой Алгебры некоторые группы первокурсников обучаются по экспериментальной методике. Студенты случайным образом были поделены на "традиционную" и "экспериментальную" группу примерно одинакового объема, и в конце обучения все студенты сдавали одинаковый экзамен. Ниже приведены результаты оценок на экзамене. Каков эффект от экспериментального обучения?

Традиционная группа	Экспериментальная группа
[illegible]	[illegible]
s = 12.3	s = 6.2
[illegible]	[illegible]

4.17. [illegible]
Вы — губернатор штата и должны решить, какое из полицейских управлений четырех городов станет победителем ежегодного конкурса по эффективности. Эффективность функционирования каждого полицейского управления выра[illegible]

Полицейское управление			
A	B	C	D
[illegible]			

Использование SPSS for Windows для расчета мер разброса

[illegible]

Демонстрационный пример с использованием SPSS 4.1. Расчет размаха и стандартного отклонения

[illegible] статистик, рассмотренных в этой главе, можно рассчитать с помощью процедур Frequencies (Частоты) или Descriptives (Описательные), с которы[illegible] *educ* (HIGHEST YEAR OF SCHOOL COMPLETED) — число лет, затраченных на обучение, *tvhours* (HOURS PER DAY WATCHING TV) — число [illegible] за просмотром телевизора. Эти три переменные мы уже использовали в демонстрационном примере SPSS 3.2.

Из главного меню нажмите Analyze (Анализ), Descriptive Statistics (Описательные [illegible]) Descriptives [illegible]. В этом окне в списке переменных слева найдите переменные *age*, *educ* и *tvhours* и переместите их в поле Variable(s) справа (для этого подведите курсор на имя переменной и нажмите кнопку с изображением стрелки). Щелкните OK, после чего появится окно результатов, аналогичное окну, которое мы анализировали в демонстрационном примере SPSS 3.2. Но теперь нас будут интересовать размах и стандартное отклонение, а не меры центра распределения.

Описательные статистики

	N	Минимум	Максимум	Среднее	Стандартное отклонение
Возраст респондента	1385	18	89	44.94	17.06
[illegible] лет, затраченных [illegible] обучение	1381	0	20	13.37	2.86
Число часов в день, проводимых за просмотром телевизора	1134	0	21	2.86	2.20

[illegible] наблюдений (объект, у [illegible] хотя бы одной переменной [illegible]

В столбце (ст. отклонение) приведено стандартное отклонение для каждой переменной. А размах можно рассчитать самим с помощью столбцов Минимум и [illegible] шему респонденту 18 лет, [illegible] менной *educ* равняется 2.86, а диапазон измен[illegible] те респонденты [illegible] чем стандартное отклонение [illegible] ное отклонение несет больше смысла, когда [illegible] это на примере. Пусть нас интересует [illegible]

или меняется с течением времени. Окно представления результатов, полученное в [illegible] Descriptives для данных 1998 г., покажет нам, что в 1998 г. [illegible] смотрели телевизор 2,86 часа в день, со стандартным отклонением [illegible] 2,20. Пусть мы из выборки [illegible] телевизор 3,70 часа в день, со стандартным отклонением [illegible]. Можно сделать вывод, что через 20 лет время просмотра телевизора в среднем уменьшилось, но американцы стали более неоднородны по этой переменной.

Демонстрационный пример с использованием SPSS 4.2. Использование команды COMPUTE для создания "шкалы отношения к абортам"

В SPSS есть несколько похожих процедур, позволяющих преобразовывать переменные и управлять ими. Так, в упражнениях в конце главы 2 мы описали команду Recode (Перекодировать) и воспользовались ею для изменения значений ассоциированных с переменной. А в этой демонстрации мы воспользуемся командой Compute (Вычислить), чтобы создать новые переменные и объединенные шкалы.

Давайте начнем с рассмотрения двух переменных — *abany* и *abhlth* — из GSS 1998 г., которые оценивают отношение респондента к абортам. Каждая из этих переменных описывает свою ситуацию и просит респондента сообщить, допускает ли он аборт в этой ситуации — независимо от другой переменной. Так, *abany* является ответом на вопрос, возможен ли аборт *по любой причине*: *abhlth* — [illegible] возможен ли легальный аборт при условии, что здоровье женщины в опасности. Так как указанные две ситуации различны, то переменные *abany* и *abhlth* следует анализировать по отдельности. Но давайте предположим, что нам нужно создать объединенную шкалу, измеряющую *общие* ощущения респондента по поводу абортов.

[illegible] последовательны выступает за *аборты*, т.е. ответил "да" [illegible] GSS [illegible] новой переменной 3 появится в том случае, если респондент [illegible] "да" на один вопрос и "нет" на другой вопрос [illegible]

Если ответ на вопрос abany таков	и	ответ на вопрос abhlth таков,	то	значение переменной abscale составит
[illegible]		За		2 (За аборты)
[illegible]		[illegible]		[illegible]
[illegible]		[illegible]		[illegible]
[illegible]		[illegible]		4 (Против абортов)

[illegible] *abscale* выражает общие ощущения респондента по [illegible] Так мы создали *abscale*, ее значения можно будет анали-

зировать, преобразовывать и управлять ими аналогично любой реальной переменной GSS, имеющейся в исходном файле данных.

Чтобы вызвать команду Compute, щелкните **Transform** (Преобразование) из главного меню и затем **Compute** (Вычислить). Появится окно с названием "Compute Variable" (Расчет переменной) [illegible] **Numeric Expression** [illegible] изображением стрелки, которая находится между списком переменных и полем **Numeric Expression**). Теперь щелкните по знаку плюс (+) панели калькулятора окна "Compute Variable", расположенной ниже поля **Numeric Expression**, — и рядом с *abany* появится плюс. Наконец, выберите в списке переменных *abhlth* и переместите эту переменную в поле **Numeric Expression** (для этого щелкните по кнопке с изображением стрелки).

Теперь в поле **Numeric Expression** у вас должно быть следующее:

abany + abhlth

Щелкните **OK**, в результате чего будет создана и добавлена к файлу данных новая переменная *abscale*. Если вы хотите постоянно хранить эту переменную, то сохраните файл данных (выбрав **Save** из меню **File**). Помните, что если вы пользуетесь *SPSS student version*, то объем файла данных ограничен 50-ю переменными.

Теперь у вас есть три переменные, измеряющие отношение респондента к абортам — две переменные, относящиеся к конкретным ситуациям, и третья, обобщающая. Всегда старайтесь проверить таблицу частот для новых переменных, полученных путем расчета или перекодировки — чтобы убедиться в том, вы не ошиблись при расчете этих переменных. Чтобы получить таблицы частот для переменных *abany*, *abhlth abscale*, воспользуйтесь процедурой **Frequencies** (Частоты). Для этого щелкните **Analyze**, **Descriptive Statistics** и затем **Frequencies**, и в появившемся окне выберите сразу три эти переменные, щелкните по кнопке со стрелкой и затем по **OK**. Вы должны получить такие результаты:

Статистика

		Аборты по желанию женщины по любой причине	В случае, если здоровье женщины в опасности	ABSCALE
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Аборты по желанию женщины по любой причине

		Частота	Процент	Процент действительных	Кумулятивный процент
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	35	[illegible]	[illegible]	[illegible]

[illegible]

	Частота	Процент	Процент действительных	Кумулятивный процент
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

В случае серьезной опасности здоровью женщины

	Частота	Процент	Процент действительных	Кумулятивный процент
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

ABSCALE

	Частота	Процент	Процент действительных	Кумулятивный процент
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Заметим, что [illegible] женщине право на аборт в ситуации, когда [illegible] не одобряют аборты по желанию [illegible]. Посмотрев на значения обобщенной переменной [illegible] в ряде ситуаций (значение 2) [illegible] и что менее 12% не [illegible] ни в одной ситуации (значение 4). [illegible]

[illegible] вопросы GSS об абортах [illegible] респонденту не задаются [illegible] пропущенных наблюдений получены [illegible] GSS, в котором среди [illegible] двух вопросов

[illegible] пропущены. [illegible] т. е. наблюдения, в которых [illegible]

Упражнения

4.1 С помощью команды **Descriptives** (в SPSS) или **Univariate** (в MicroCase) [illegible] (пол респондента), *papres80* (престиж профессии отца респондента).

а. [illegible]

б. [illegible] Является ли респондент более образованным, чем [illegible] по переменной [illegible]

4.2 С помощью команды **Compute** (в SPSS) или **Sum Recode** (в MicroCase) создайте [illegible] переменных *[illegible]* и *[illegible]*. Рассчитайте [illegible] новой переменной, а также для [illegible] выводы [illegible] таблицами [illegible] опишите распределения переменных [illegible]

5 Кривая нормального распределения

Цели изучения После изучения данной главы вы смож[illegible]

1. Давать определение и объяснить понятие кривой но[illegible]ного распределения
2. [illegible] нии и кривой нормального распределения (приложение А) для нахождения областей до, после и между точками на к[illegible]
3. Выражать области под кривой в терминах ве[illegible]

5.1. Введение

Кривая нормального распределения [illegible] тистике. Как мы увидим в части II, она является центральной для теории, лежащей в [illegible]. Также, в сочетании со стандартным отклонением, кривая нормального распределения может быть использована для построения [illegible] эмпирических распределений. Данная глава за[illegible] вершит наше рассмотрение статистики [illegible]

[illegible]

[illegible] рмального распределения [illegible] ная ось) распределения, изме[illegible] ния от среднего арифметического распре[illegible] площади под кри[illegible]. Другими словами, если

[illegible] природу, лежащую в основе распределения, расстояни[illegible] любой точки до среднего (измеренного в количестве стандартных отклонени[illegible]) [illegible] площадь.

Для того чтобы проиллюстрировать все вышеизложенное на рис. 5.1 и 5.2 представлено два гипотетических распределения коэффициентов IQ, одно — для гру[illegible] [illegible] распределением (либо практически нормальным) со следующими параметрами:

Мужчины	Женщины
$\bar{X}$ = [illegible]	[illegible]
s = [illegible]	[illegible]
N = 1000	N = 1000

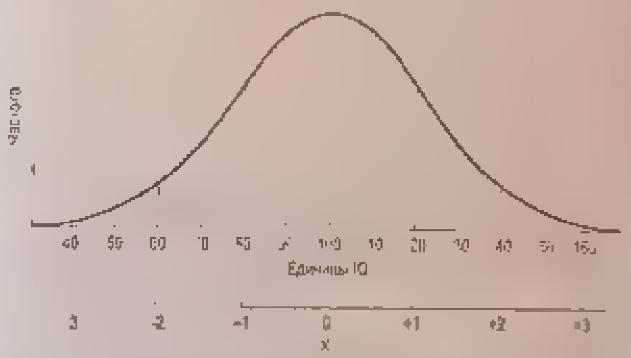


Рис. 5.1. *Коэффициенты IQ для выборки [illegible]*

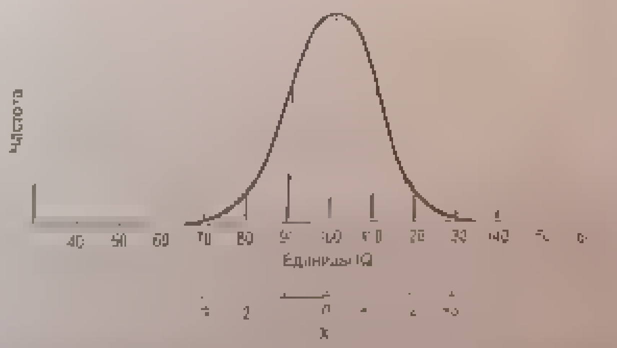


Рис. 5.2. *Коэффициенты IQ для выборки женщин*

[illegible] на горизонтальной оси (или абсц[illegible]) [illegible] стандартного отклонения, и областями под кр[illegible]

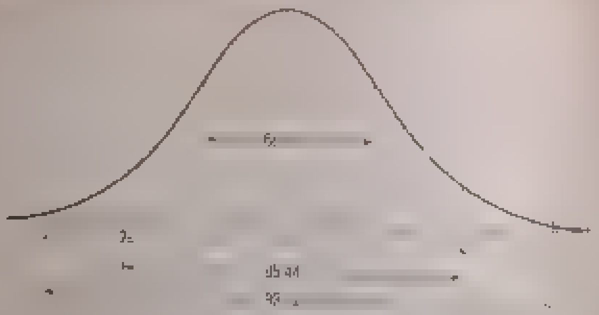


Рис 5.1

5.2. Расчет стандартизованных значений (Z-значений)

Формула 5.1

5.3 Таблица кривой нормального распределения

[illegible] находится между Z-значением, равным 1,00, и среднего", либо [illegible] под кривой находится между значением, равным 1,00, и среднего 34,13% общей площади находится между значением, равным 1,00, и среднего [illegible]

Третий столбец представляет *Об*[illegible] Z [illegible] [illegible] за отрица[illegible] [illegible] ваться в тех случаях, когда мы хотим найт[illegible] Z-значениями (его применение будет объяснено в разделе 5.4).

Для экономии места таблица содерж[illegible] Z [illegible] [illegible] кривой нормального распределения [illegible] относящиеся к любому отрицательному значени[illegible] [illegible] Например [illegible] Z [illegible] нием, равным −1,00 и среднего [illegible] 34 [illegible] которую мы нашли ранее для значения, равного +1,00. Тем не менее, как мы [illegible] неоднократно будем показывать их же, знак Z-значения чрезвычайно важен и должен быть точно отмечен.

Для того чтобы попрактиковаться в использовании приложения А для описания областей под эмпирической кривой нормального распределения [illegible] ния их Z-значения и области, приведенные ниже для распределения IQ мужчин. Для каждого значения IQ с использованием формулы 5.1 рассчитывается соответствующее Z-значение. Затем для того, чтобы найти области между Z-значением и средним используется приложение А. ($\bar{X}$ = 100, s = 20 везде)

IQ	Z [illegible]	Области между Z-значением и средним
110	+0,50	19,15%
125	+1,25	39,44%
133	+1,65	45,05%
[illegible]	+1,90	47,13%

Та же методика применяется, когда Z-зна[illegible] чение является отрицательной величиной (такое происходит в тех случаях, когда исходное значение находится ниже среднего).

[illegible]	[illegible]	Области между Z-значением и средним
93	−0,35	13,68%
85	−0,75	27,34%
67	−1,65	45,05%
62	−1,90	47,13%

[illegible]

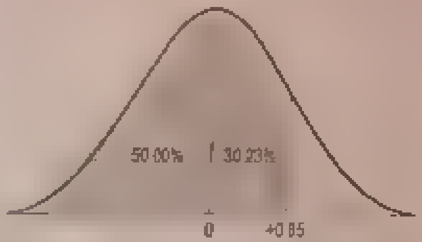
5.4. Расчет общей площади перед и после значения

До этого мо[illegible] мы видели, как таблица кривой нормального распределения [illegible] Z [illegible] [illegible] Z-значению +0,85

$$Z = [illegible]$$

Положительный знак Z [illegible] что значение должно находиться за (справа) от среднего. Чтобы найти область ниже положительного Z-значения, область между значением и средним должна быть прибавлена к области ниже среднего. Поскольку кривая нормал[illegible] на медиане, а область перед средним, таким образом, будет равна 50%. Внимательно изучите рис. 5.4. Нас интересует заштрихованная область.

Рис. 5.4
Нахождение области перед положительным Z-значением



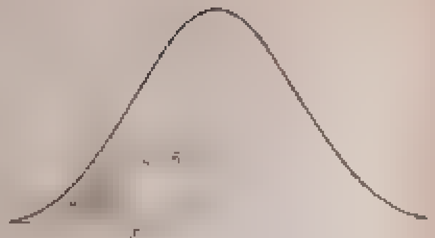
[illegible]

$$Z = [illegible]$$

[illegible]

[illegible] как 0,0885, что [illegible] получил больше баллов, чем 8,85% тестируемой группы.

Рис. 5.5
Нахождение области перед [illegible]



В приведенном выше примере мы используем определенную методику для того, чтобы найти область перед значением. По существу, та же методика используется для нахождения области после значения. Если нам, например, необходимо определить область после 108 балла IQ, мы сначала переведем баллы в *Z*-значения

$$Z = \frac{X - \bar{X}}{s} = \frac{108 - 100}{20} = \frac{8}{20}$$

[illegible] к приложению А. Заштрихованная область на рис. 5.6 показывает [illegible] которая нас интересует. Область за положительным значением можно найти в столбце *Область за Z*, и, в данном случае, область равна 0,3446 или 34,46%.

Рис. 5.6
[illegible]

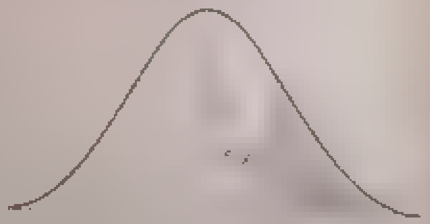


Таблица 5.1. Расчет областей перед и за положительными и отрицательными *Z*-значениями

	Когда *Z*-значение	
Для нахождения области [illegible]	[illegible]	отрицательно
[illegible]	[illegible]	прибавьте к области в столбце B 0,5000 или 50,00%
[illegible]	[illegible]	в столбце C

[illegible] ики обобщены в таблице 5.1. Для того чтобы найти общую площадь [illegible] *Z*-значением или за отрицательным *Z*-значением, двигайтесь [illegible] нас значение. Область [illegible] за *Z*. Для того, чтобы найти [illegible] или после отрицательного зна-

[illegible] *Z* [illegible] кривую и заштрихуйте интересующие [illegible] для областей перед и после *Z* значения [illegible]

5.5 Определение области между двумя значениями

[illegible] 1,2 баллами IQ, мы должны перевести оба значения в *Z*-значения, найти область между каждым значением и средним в приложении А и сложить эти две области. 93 балла IQ преобразуются в *Z*-значение –0,35:

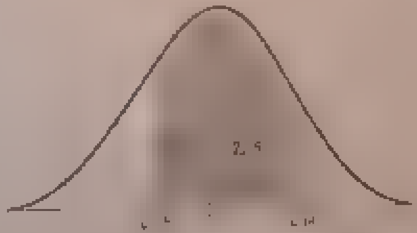
$$Z = \frac{X - \bar{X}}{s} = [illegible]$$

[illegible] баллов IQ [illegible]

$$Z = \frac{X - \bar{X}}{s} = \frac{112 - [illegible]}{20} = [illegible]$$

Оба значения показаны на рис. 5.7. Нас интересует вся заштрихованная область. Общая площадь между этими двумя значениями равна 13,68%+22,57% или 36,25%. Таким образом, 36,25% общей площади (или около 363 из 1000 случаев) находится между 93 и 112 баллами IQ.

Рис. 5.7
Нахождение области между двумя значениями



Практический пример 5.1

Вы только что получили свои результаты [illegible] набрали 7[illegible] баллов [illegible] вы знаете, что [illegible] стандартное отклонение 5 [illegible] результат [illegible]

[illegible] что результаты теста распределены нормально [illegible] и найти область перед или после вашего [illegible] эквивалент вашего исходного балла будет [illegible]

$$Z = [illegible]$$

$$Z = [illegible]$$

$$Z = [illegible]$$

$$Z = [illegible]$$

Обратившись к приложению А, мы узнаем, что *Область между средним и Z* составляет 0,4861, что также может быть выражено как 48,61%. Поскольку это положительное Z-значение, мы должны прибавить данную область к 50,00% для того чтобы найти общую область перед значением. Ваш балл выше, чем (48,61+50,00) или 98,61% всех результатов теста. Вы неплохо справились.

Когда интересующие нас значения находятся по одну сторону от среднего, то для определения области между ними необходимо придерживаться другой методики. [illegible], что преобразовали бы эти баллы в Z-значения

$$Z = [illegible] = \frac{[illegible]}{20} = [illegible]$$

$$Z = \frac{X_2 - \bar{X}}{[illegible]} = \frac{121 - 100}{[illegible]} = \frac{21}{[illegible]} = [illegible]$$

[illegible] нас интересует заштрихованная область. Для [illegible] области между двумя значениями по одну сторону от среднего, [illegible] области между каждым значением и средним (приведено в приложении А) [illegible] Таким образом, область между этими двумя значениями [illegible] или 11,09% общей площади. Необходимо следовать этой [illegible] значения находятся перед средним. Методики определения [illegible] значениями обобщены в таблице 5.2 [illegible] *между двумя значениями, см. задачи 5.3, 5.4.* [illegible]

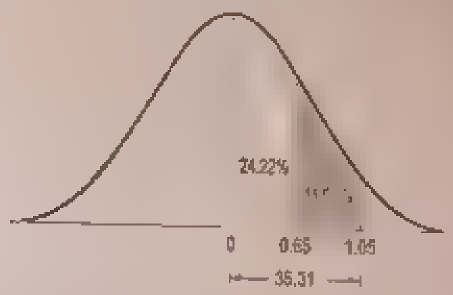


Рис. 5.8
Нахождение области между двумя значениями

Таблица 5.2 Расчет областей между двумя значениями

Ситуация:	*Методика:*
Значения находятся по ОДНУ сторону от среднего.	Найдите области между каждым значением и средним в столбце b. Отнимите меньшую область от большей.
Значения находятся по РАЗНЫЕ стороны от среднего.	Найдите области между каждым значением и средним в столбце b. Сложите две области.

5.6. Использование кривой нормального распределения для оценки вероятностей

[illegible] мы рассматривали теоретическую кривую нормального распределения как способ описания доли от общей площади перед, после и между значениями в эмпирическом распределении. Мы также увидели, что эти области могут быть преобразованы [illegible] Теоретическая кривая нормального распределения может также рассматриваться как распределение вероятностей. Кроме того, мы можем использовать свойства теоретической кривой нормального распределения [illegible] того, попадет ли в определенный диапазон наблюдение, выбранное наугад из эмпирического нормального распределения. Что же касается методики, то данные вероятности будут определяться таким же путем, как и области, которые мы уже находили. Тем не менее, прежде чем обсудить эти методики, давайте разберемся, что понимается под понятием вероятности.

Хотя мы редко об этом задумываемся, мы каждый день стараемся использовать [illegible]

[illegible]

[illegible] е составят успех, в знаменатель должен быть равен общему [illegible] в котором теоретически может возникнуть успех [illegible]

[illegible] вести дробь. Лишь одна карта отвечает нашему определ[illegible] зом, количество событий, которые составят успех, равно 1. Это число будет [illegible] ает 52 возможных события (поскольку в колоде находится [illegible] является вероятностью вытянуть червового короля с первого р[аза] [illegible] мешанной колоды карт. Наша вероятность успеха 1 из 52.

Мы можем оставить полученную выше дробь в таком виде, в [illegible] же мы можем выразить ее несколькими другими способами. Например, мы можем выразить ее как обратную дробь, путем переворота дроби, показав, что [illegible] тавить червового короля с первого раза составляют 52 1 (или 52 к 1). Мы можем выразить дробь как пропорцию, разделив числитель на знаменатель. Для нашего примера соответствующая пропорция составит 0,0192, что является долей вероятных событий, которые удовлетворят нашему определению успеха. В социологии вероятности обычно выражаются как пропорции, и в дальнейшем мы будем следовать дан[illegible]

Как это было здесь отмечено, вероятности имеют точное значение: в большом количестве событий, которые мы определили как успех, они составят определенное пропорциональное соотношение к общему количеству событий. Вероятность 0,0193 вытянуть червового короля с первого раза в действительности означает, что из тысяч вытягиваний одной карты за раз из полной колоды в 52 карты, пропорция успешных выпадений составит 0,0192. Или из каждых 10 000 вытягиваний, 192 раза [illegible] когда мы говорим, что вероятность вытянуть червового короля за один раз состав[ляет] [illegible]

Как и пропорции, вероятность находятся в промежутке от 0,00 (означает, что [illegible] не может произойти ни при каких обстоятельствах) до 1,00 (несомненный [illegible] увеличения значения [illegible] [вы]тянуть червового короля) маловероятна [illegible] нию вероятности с нашим знанием теоретичес[illegible] определения, мы можем [illegible]

[illegible] межутке, а знаменатель равен общему колич[illegible] пирическое распределение нормально, мы можем пропустить данный шаг, поскольку [illegible]

[illegible]

[illegible] ной в данном разделе, является идея о том, что площади в таблице кривой нормального распределения могут также рассматриваться как вероятности.

Наконец, необходимо обратить внимание на очень важное обстоятельство, связанное с вероятностями и кривой нормального распределения. Вероятность того, что любое наблюдение, выбранное наугад из нормального распределения, будет иметь величину, очень близкую к значению среднего, очень высока. Форма кривой нормального распределения такова, что большинство случаев сгруппированы вокруг среднего и их частота снижается по мере того, как мы удаляемся (влево или вправо) от среднего значения. В действительности, принимая во внимание то, что мы знаем о кривой нормального распределения, вероятность того, что случайно выбранное наблюдение будет иметь балл IQ в пределах ±1 стандартное отклонение от среднего, равно 0,6826, в то время как вероятность того, что наблюдение будет иметь значение более чем в 3 стандартных отклонениях от среднего, равна 0,0028 (1,000 - 0,9972). Таким образом, если мы случайно выберем некоторое количество случаев из нормального распределения, мы выберем много наблюдений, чьи значения близки к [illegible] меньше среднего. *(Для получения практических навыков использования таблиц кривой нормального распределения* [illegible]

Заключение

1. Кривая нормального распределения в сочетании со средним и стандартным от[illegible] ходу. Данная глава является также важнейшим теоретическим базисом для [illegible] 11.
2. [illegible] ные баллы должны быть [illegible] вленные значения. Стандартизованные значения дают [illegible]

[illegible] теоретической кривой нормального распределения (приложение А).

3. Мы [illegible] три области применения теоретической кривой нормального распределения [illegible] частей. Эта последняя возможность [illegible] распределения особенно важна, поскольку статистические выводы [illegible] с оценкой вероятности различных событий, который [illegible] бом, очень похожим на тот, который представлен в разделе 5.6.

Формулы

Z-значения [illegible] $Z = \frac{X_i - \bar{X}}{s}$

Словарь

Кривая нормального распределения. [illegible] Симметричная, имеет одну вершину и форму колокола. Стандартная кривая нормального распределения всегда имеет среднее арифметическое, равное 0, и стандартное отклонение, равное [illegible]

Таблица нормального распределения. Приложение А, подробное описание [illegible] ти между Z-значением и средним любого стандартизированного нормального распределения.

Z-значения. Стандартизованные значения; способ, с помощью которого выражаются значения после того, как они были приведены к виду теоретической кривой нормального распределения.

Мультимедийные ресурсы

The Wadsworth Sociology Resource Center [illegible] Центра социологических ресурсов Вадсворта)

[illegible]

Посетите сайт шестого издания *Statistics: A Tool for Social Research* (статистика — инструмент [illegible] учебным материалам

- [illegible]
- Обучающие карточки
- Дополнительные задачи к главам
- [illegible]
- [illegible]
- [illegible]

- Примеры и упражнения для использования пакетов статистической обработки данных MicroCase и SPSS;
- Функциональные диаграммы "Find the test"
- Проверка гипотез для переменных, измеряемых по порядковой шкале

Задачи

5.1. [illegible] нормально распределены и имеют среднее, равное 10, и стандартное отклонение — равное 3. Для каждого приведенного ниже значения определите Z-значение и процент области перед и после этого значения.

X_i	Z-значение	% области после	% области перед
5			
6			
7			
8			
9			
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			
8			

5.2. Предположим, что распределение баллов, полученных на вступительных экзаменах в вуз, является нормальным. Среднее равно 500, а стандартное отклонение — 100. Для каждого приведенного ниже значения определите соответствующее Z-значение, процент области перед значением и процент области после значения.

X_i	Z-значение	% области после	% области перед
65[illegible]			
[illegible]00			
[illegible]			
586			
[illegible]			
5[illegible]6			
62[illegible]			
495			
[illegible]			
398			

5.3. В [illegible] классе провели итоговый экзамен [illegible] Среднее [illegible] и стандартное отклонение [illegible] Как [illegible]

а) между 75 и 85
б) между 80 и 85
в) после 80
г) [illegible]
д) между 60 и 70
е) между 75 и 70
ж) перед [illegible]
з) перед 77
и) перед 80
к) перед 95

5.4. Для нормального распределения, в котором среднее равно 50, а стандартное отклонение — 10, найдите процент площади

а) между 40 и 47
б) после 47
в) перед 53
г) между 35 и 65
д) после [illegible]
е) перед 31 и после 69
ж) [illegible]
з) между 32 и 47

5.5 В колледже Святой Алгебры 200 первокурсников, [illegible] которому средний балл составил 72 [illegible] стандартное отклонение [illegible] приведены оценки 10 студентов. Преобразуйте [illegible] количество людей, которые [illegible] каждого из 10 студентов. *Подсказка: умножьте соответствующее [illegible] на N и округлите результат.*

X	Z значение	Количество людей после	Количество людей перед
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			

5.6. Если распределение оценок по тесту нормальное, среднее равно [illegible], а стандартное отклонение — 11, то какой процент области находится:

а) перед 60
б) перед 70
в) перед 80
г) перед 90
д) между 60 и 65
е) между 65 и 79
ж) между 70 и [illegible]
з) между 80 и 90
и) после 99
к) после 89
л) после 75
м) после 65

5.7 SOC

Шкала, измеряющая предубеждение респондентов была [illegible] большой выборки. Распределение баллов по этой шкале приблизительно нормальное, среднее равно 31, а стандартное отклонение — 5. Какой процент выборки получил:

а) менее 20 баллов
б) меньше 40 баллов
в) между 30 и 40 баллами
г) между 35 и 45 баллами
д) больше 25 баллов.
е) больше 35 баллов.

5.8. CJ

Среднее количество краж со взломом составляет 3.1 в год, а стандартное отклонение — 50. Какова вероятность того, что в следующем году количество [illegible]

а) менее 250:
б) менее 300
в) более 350
г) более 400
д) между 250 и 350:
е) между 300 и 350;
ж) между 350 и 375.

5.9. Какова вероятность того, что по тесту по математике, для которого среднее составляет 59, а стандартное отклонение — 4, случайно выбранный студент получит [illegible]

а) между 55 и 65;

б) между 60 и 65;

в) выше 65;

г) между 60 и 50;

д) между 55 и 50;

е) ниже 55.

5.10. SOC

Если для шкалы, приведенной в задаче 5.7, 40 и более баллов рассматривается как *сильно предубежден*, то какова вероятность того, что случайно выбран[illegible] человек получит такое количество баллов?

5.11. Полиция города Шинбон, штат Канзас, дает кандидатам вступительный экзамен и принимает только тех, кто попадет в 15% по количеству набранных баллов по данному тесту. Если средний балл в этом году составил 87, а стандартное отклонение — 8, примут ли человека, набравшего 11 баллов?

5.12. [illegible] экзамена на должность социального работника и [illegible] консультанта по трудоустройству [illegible] информацию о тестах и о своих результатах. С каким из тестов вы справились лучше?

Социальный работник	Консультант по трудоустройству
$\overline{X}$ = 118	$\overline{X}$ = 27
s = 7	s = 3
Ваш результат = [illegible]	Ваш результат = 29

5.13. В распределении [illegible], где среднее равно 35, а стандартное отклонение — 4, вероятность какого события более высока: что случайно выбран[illegible] значение будет находиться между 29 и [illegible] или что случайно выбранное значение будет находиться между 40 и 42?

5.14. Для того чтобы стать членом почетного общества, результат студента по GPA (средний балл) должен входить в [illegible]% лучших в школе. Если средний GPA составляет 2,78, а стандартное отклонение — 0,33, какие из приведенных ниже GPA будут подходящими?

3,20; 3,21; 3,25; 3,30; 3,35

SPSS for Windows

Использование SPSS for Windows для преобразования исходных значений в стандартизированные

Демонстрационный пример с использованием SPSS 5.1. Расчет стандартизованных значений

Программа Descriptives, которая была представлена в конце главы 4, также может быть использована для расчета Z-значений (стандартизованных значений) [illegible]

[illegible] Эти Z-значения затем [illegible] SPSS [illegible] новую переменную, содержащую преобразованные [illegible] букву Z и первые 7 букв названия [illegible] переменной.

В [illegible] SPSS рассчитывает Z-значения для *возраста*. В [illegible] выберите набор данных [illegible] и нажмите Analyze, Descriptive Statistics и Descriptives (Анализ, Описательные статистики, [illegible]). [illegible] Variable(s) (Переменные) [illegible] Standard [illegible] as variables [illegible] Descriptives SPSS [illegible] Variable(s). Нажмите OK, и SPSS [illegible] *age* [illegible] *zage* [illegible] Descriptives [illegible] *zage* [illegible] Variable(s) [illegible] *age*, нажмите OK и [illegible]

Показатели описательной статистики

	N	Минимум	Максимум	Среднее	Стандартное отклонение
Возраст респондента	1385	18	89	44,94	17,08
Z-значение: возраст респондента	1385	[illegible]	[illegible]	2,86E-15	[illegible]
Значимое количество (в обратном порядке)	1385				

Обратите внимание, что средняя для *zage* дана как 2,86E-15. Это форма экспоненциального представления чисел, которое можно интерпретировать, переставив запятую десятичной дроби налево на столько позиций, сколько определено числом, следующим за E (-0,00000000000000286). Другими словами, как и любой другой набор Z-значений, *zage* имеет среднее, равное нулю, и стандартное отклонение, равное [illegible]. Новую переменную *zage* можно рассматривать просто как любую другую переменную и использовать в любых других операциях, выполняемых с помощью SPSS.

Если вы захотите просмотреть значения *zage* [illegible] Case Summaries [illegible] Analyze, Reports (Анализ, Отчеты), а затем Case Summaries. [illegible] *zage* [illegible] Variable(s). Найдите опцию Limit cases to first [illegible] для задания количества наблюдений, включенных в результат. Для данного упражнения зададим лимит на уровне 10 [illegible] OK [illegible] посмотра результатов будет получен следующий отчет:

Сводка наблюдений

	Возраст респондентов	Z-значение: Возраст респондентов
1	6[illegible]	1.68 6[illegible]
2	17	-1,05030
3	[illegible]	4[illegible] 6[illegible]
4	[illegible]	[illegible].58 [illegible]
5	43	-0.11351
6	29	-0.93330
7	39	-3.4770
8	45	0.00359
9	29	0.93320
10	41	-0.23061
[illegible]	[illegible]	0

Ограничено до первых 10 наблюдений

Просмотрите список значений и обратите внимание, что значения, которые близки по величине к среднему *age* (44,94), также очень близки к среднему *zage* (0,00), и чем дальше значение от 44,94, тем больше величина Z-значения. Также обратите внимание, что, конечно же, значения перед средним (меньшие 44,94) имеют отрицательный знак, а значения после среднего (большие 44,94) имеют положительный знак.

Упражнения

5.1. Рассчитайте Z-значения для переменных *prestg80*, *papres80* и *tvhours*.

а. (SPSS) Используйте подпрограмму **Case Summaries** для получения стандартизованных и исходных значений для каждой переменной для 10 наблюдений.

б. (MicroCase) Используйте команду **Univariate** для рас[illegible] показ[illegible] тельной статистики и частотные распреде[illegible] для [illegible] ре[illegible] менных.

5.2. [illegible] **Graphs** для получения простых линейных [illegible] и *zage*. Как близки эти графики по форме к глад[illegible] кривой нормального распределения?

Итоговые упражнения к части I

1. До какой степени люди считают религию одной из важнейших вещей в жизни? 15 человек ответили на ряд вопросов, [illegible] следующие:

   Ваше вероисповедание?

   1. Протестант
   2. Католик
   3. Иудей
   4. Атеист
   5. Другое (мусульманин, индуизм, др.)

2. По шкале от 1 до 10 (где 10 — наибольшее) насколько сильна ваша вера?
3. Сколько раз в день вы молитесь?
4. Когда дела идут не слишком хорошо, моя вера является для меня главным источником спокойствия:

   1. Полностью согласен
   2. Частично согласен
   3. Частично согласен, частично не согласен
   4. Частично не согласен
   5. Полностью не согласен

5. Сколько вам лет?

Наблюдение	*Религия*	*Сила*	*Молитва*	*Спокойствие*	*Возраст*
1	1	2	0	2	50
2	2	8	1	1	67
3	1	8	3	1	45
4	1	6	0	1	43
5	3	9	3	1	34
6	3	8	0	3	18
7	4	0	0	5	52
8	5	9	6	1	37
9	1	5	0	2	54
10	2	6	2		55
11	2	3	0	5	33
12	1	6	1	3	45
13	1	8	2	2	37
14	1	7	2	1	40
15	2	8	1	1	25

Для каждой переменной рассчитайте частотные распределения и соответствующие показатели центра распределения и дисперсии. Напишите предложение, характеризующее каждую переменную.

2. [illegible]

В. [illegible]

В. [illegible]

1. Мужской

2. Женский

Браки между людьми разных расовых групп неудачны и должны быть за[illegible] [illegible]

1. Полностью согласен
2. [illegible]
3. Не [illegible]
4. Не [illegible]
5. Полностью не согласен

4. Сколько лет вы проучились?

5. Какая из нижеперечисленных категорий наилучшим образом описывает [illegible] в котором вы выросли?

1. Большой город
2. Город среднего размера
3. Пригород
4. Маленький город
5. Сельская местность

6. Ваше семейное положение

1. Женат/замужем
2. Разведен/разведена
3. Вдовец/вдова
4. Никогда не был женат/замужем

Ниже воспроизведены ответы 20 респондентов

Наблюдение	Возраст	Пол	Отношение к межрасовым отношениям	Годы обучения	Местность	Семейное положение
[illegible]			5			
2	25	2	4	[illegible]		4
3	55	2	[illegible]	6	2	
[illegible]	4[illegible]		3	9	2	1
[illegible]	2[illegible]	2		2	3	1
[illegible]	[illegible]			10	1	1
[illegible]	[illegible]	2	4	5	5	
				10	[illegible]	2

Наблюдение	Возраст	Пол	Отношение к межрасовым отношениям	Годы обучения	Местность	Семейное положение
8	[illegible]			2	[illegible]	
9	3[illegible]			2		
[illegible]	42	2	4	5		4
11	[illegible]	[illegible]	4	[illegible]	[illegible]	
[illegible]	[illegible]	2	2	[illegible]	4	
[illegible]3		1	3	9		2
[illegible]	[illegible]	1		5	1	
[illegible]5	[illegible]	1		[illegible]		[illegible]
[illegible]6	[illegible]5			[illegible]	4	
[illegible]	[illegible]	2	[illegible]	5	5	
[illegible]8	42	2				4
[illegible]9	[illegible]	[illegible]				[illegible]
20	2[illegible]	[illegible]				[illegible]

а. [illegible]

б. Для всей 1000 респондентов средний возраст составил 34.70, а стандартное отклонение [illegible] [illegible] [illegible] те процент области перед (младше чем) каждым респондентом

3. [illegible]

Сколько детей вы имели (значения являются фактическим количеством)?

[illegible]

[illegible]

4. [illegible]

3. Раса

1. Белая
2. Черная
3. Другая

Тара-ра, тара ра.

... способы контроля рождаемости должны быть доступны подросткам

6. Ваше религиозная принадлежность

1. Протестант
2. Католик
3. Иудей
4. Атеист

Количество ...	Образование	Раса	Контроль рождаемости	Телевизор	Религия

Часть II

СТАТИСТИЧЕСКИЕ ВЫВОДЫ

6 Введение в статистические выводы. Формирование выборок и распределение выборочных статистик[1]

Цели изучения

К концу этой главы вы сможете

1. Объяснить назначение ста[illegible]их выводов как метода распростра[illegible]ение результат[illegible] [illegible]борки, на генеральную совокупность
2. Дать определения и пояснить на примерах основн[illegible] [illegible]ды формирования случайных выборок
3. Пояснить и дать определения таких ключевых понятий: генеральная совокупность, выборка, парамет[illegible], статистика, репрезентативность, метод равновероятностного отбора (МРВО)
4. Различать между собой такие понятия: распределение выборочных статистик, выборка, генеральная совоку[illegible]
5. Пояснить смысл двух теорем, представленных в этой главе

6.1. Введение

[illegible]

[illegible]ентов, которые сильно различаются [illegible]

[illegible]

возможно [illegible] мы не сможем опросить каждого члена интересующей нас [illegible] которая [illegible] почти [illegible] миллионов человек). В действительности, даже для тех [illegible] больших групп [illegible] [illegible] материальное обеспечение [illegible] членах данной группы не представляется возможным.

[illegible] выводы часто называются параметрами) на основе анализа полученных выборок. В этой книге мы рассмотрим два варианта применения статистических выводов. В процедурах [illegible] генеральной совокупности, на основании того, что известно о имеющейся выборке. А при проверке гипотез, описанных в главах 8–11, проверяется их справедливость для генеральной совокупности путем их проверки на выборочных данных. В этой главе [illegible] рассмотрим формирование выборок [illegible] выборки) и затем введем ключевое понятие статистических выводов: распределение выборочных статистик.

6.2. Методы формирования вероятностных выборок

Социологи разработали большое количество методов формирования выборок и генеральных совокупностей. В этой главе мы приведем обзор основных процедур формирования вероятностной выборки — единственного типа выборки, к которой можно в полной мере применить методы получения статистических выводов о генеральной совокупности. [illegible] му в следующих главах я часто буду его использовать. Однако термин "вероятностная выборка" все же предпочтительнее, т.к. в повседневной речи под случайным мы [illegible] виду непредсказуемость наших действий. Но, как вы узнаете из этой главы, вероятностные выборки формируются [illegible] [illegible] чайным, но к сформированной вами выборке нельзя будет примени[illegible] статистических выводов.

Перед тем, как расс[illegible] ые выборки, хочу обратить ваше вни[illegible] [illegible] ных целей (например, для поиска идей [illegible]) роль, перед тем, как начать выполнять более [illegible]

выборки обычно достаточно просто и недорого. Главный недостаток таких выборок состоит в том, что их результаты мы не имеем права обобщать на других людей. [illegible]

[illegible] совершенно правильной монеты иногда [illegible] МРВО-выборка может иногда давать нетипичные [illegible] Большим достоинством методов [illegible]

6.3 Методы формирования выборок по принципу МРВО

[illegible] разновидность метода формирования [illegible] для получения простой случайной вы-

борки [illegible]

[illegible] идентификационный номер попадается больше одного раза, игнорируйте эти повторения.

[illegible] систематической выборки [illegible]

Заметим, что при формировании выборок по систематической последовательности [illegible] строгого соответствия критерию МРВО, т.е. после того, как отобран первый [illegible] является одинаковой [illegible]

[illegible] списка, то вы можете столкнуться с трудностями из-за того, что некоторые харак- [illegible] (например, ирландские фамилии часто начинаются с "О"). Следовательно, при фор- [illegible] бое внимание природе списка вашей генеральной совокупности, а также вашему методу отбора.

[illegible] расслоенная (стратифицированная) выборка [illegible] вилность хороша тем, что гарантирует репрезентативность выборки в отношении [illegible] репрезентативной для данной генеральной совокупности.

Например, вы хотите сформировать выборку объемом 300 студентов, в которой был бы пропорционально представлен каждый [illegible] у которых профилирующий предмет — зооло- [illegible] ете генеральную совокупность [illegible]

[illegible]

До настоящего момента мы рассматривали методы формирования [illegible] простой процесс, заключающ[illegible] первичной и самой нелегкой частью работы, в составлении [illegible] что кажется на [illegible] книга, скорее всего, будет устаревшей.

Социологи придумали несколько способов борьбы с ограничениями, накладываемыми неполнотой списков. Наверное, самый важный из них — групповая (кластерная) выборка [illegible] мировать групповую выборку для вашего города так: сначала пронумеровать все [illegible] рования выборки из номеров округов. Второй этап процесса формирования групповой выборки будет заключаться в нумеровании кварталов внутри каждого отобранного ранее округа и затем, с помощью МРВО, — формировании выборки из номеров кварталов. Третий шаг может включать отбор домовладений внутри каждо- [illegible] причем вы даже не пользовались списком жителей города.

[illegible] тодами, описанными выше. Групповая выборка является менее точным представи- [illegible] элементов генеральной совокупности производится за один этап. А в случае груп- [illegible] вую выборку. В случае простой случайной [illegible] ку только один раз, а в случае групповой выборки — столько раз, [illegible] пов.

Итак, хотя мы должны относиться к выводам [illegible] точный список всего населения города может оказаться [illegible] возможно), а все, что вам нужно для формирования [illegible] та города (или список избирательных окру[illegible]

[illegible] итогом давайте вернемся к главному. Целью статистических выводов [illegible] анализа выборок, извлеченных из данной генеральной совокупнос- [illegible] МРВО [illegible] ность, но все же высока вероятность того, что МРВО-выборка будет репрезентативной для данной генеральной совокупности, из которой она извлече- [illegible]

6.4. Распределение выборочных статистик

[illegible] каким-нибудь разновидностью метода МРВО, что же мы узнали? С одной стороны, мы [illegible] почерпнуть большое количество информации из элементов нашей выборки, [illegible] мы ничего не знаем о данной генеральной совокупности. Действительно, [illegible] у нас была информация о данной генеральной совокупности, то нам бы и [illegible] понадобилась! Вспомните, что мы используем методы статистических [illegible] с целью узнать как можно больше о генеральных совокупностях, и инфор- [illegible] на основе данной выборки, важна в основном только тем, что [illegible] нам сделать обобщения, касающиеся генеральной совокупности.

[illegible] использовании методов статистических выводов мы обычно измеряем не- [illegible] ческая партия, мнения по поводу абортов), [illegible] центра распределения, 3) ка- [illegible] информации о данной перемен- [illegible] имея элементы выборки. Не менее [illegible] о переменной в генераль- [illegible] случаев (например, считается, что IQ и [illegible] нормальному распределению), нам ничего не из- [illegible] переменной в данной генеральной совокуп- [illegible] переменных данной генеральной сово- [illegible] напомнить, что если бы мы имели такую [illegible] совокупности, то статистические выводы были [illegible]

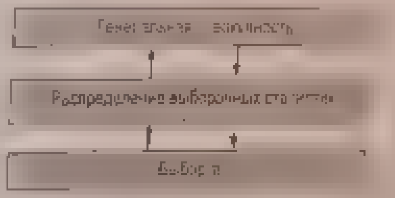
[illegible] распределение выборочных ста- [illegible]

[illegible] ристики.

[illegible] ных друг от друга распределения.

1. [illegible] известным в том смысле, что его форму, центр распределения и разброс [illegible] формация из данной выборки важна в основном тем, что позволяет исследовате- лю узнать что-то о генеральной совокупности.
2. [illegible] ким, но неизвестно нам. Единственной целью методов статистических выводов как раз является сбор информации и получение выводов, касающихся генераль- ной совокупности.
3. Распределение выборочных статистик, которое является неэмпирическим, т.е. теоретическим. Благодаря законам теории вероятности мы много знаем об этом распределении. В частности, можно получить его форму, центр распределения и разброс, следовательно, адекватно характеризовать это распределение.

Рис. 6.1. *Связь между: 1) выборкой, 2) распределением выборочных статистик и 3) генеральной совокупностью.*



[illegible]

[illegible] мы можем его построить. Здесь в качестве статистики мы [illegible] тическое. Пусть мы хотим собрать некоторую информацию [illegible] кретной общины, состоящей из 10000 человек. [illegible]

Тара-ра, тара-ра.

[illegible] из по 100 человек, которые можно изв[illegible] [illegible] ности объемом 10000 человек, а значение 27 [illegible] ных выборочных исходов.

Рис. 6.2
[illegible]

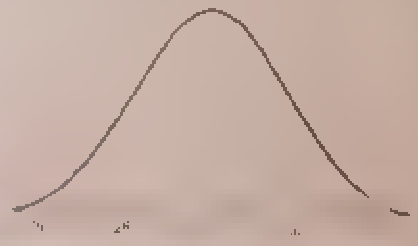
Теперь давайте заменим 100 респондентов первой выборки на других людей и извлечем вторую выборку того же объема (N = 100). Снова рассчитаем средний возраст. Пусть среднее арифметическое, рассчитанное по второй выборке, составляет 30 лет; отметим этот выборочный исход на рис. 6.2. Наша вторая выборка — еще один представитель из огромного количества возможных комбинаций по 100 человек, которые можно извлечь из данной генеральной совокупности объемом 10000 человек, а значение 30 — еще одно из множества возможных выборочных исходов. За[illegible] [illegible] третью выборку [illegible] и отметим его на рис. 6.1. Затем извлечем трет[illegible] и так будем продолжать огромное число раз, рассчитывая среднее каждой выборки и отмечая его на графике. Теперь попытайтесь представить себе, как будет выглядеть рис. 6.2 после того, как мы соберем тысячи индивидуальных выборок и рассчитаем среднее каждой выборки. Какая форма, среднее и ст[illegible] этого распределения выборочных средних после [illegible] ных комбинаций по 100 респондентов из гене[illegible] 10000 человек?

С одной стороны, мы знаем, что каждая выборка будет, по крайней мере, слегка [illegible] от каждой [illegible] мы исключаем повторный опрос в точности тех же 100 человек. Следовательно, т.к. каждая выборка соответствует уни[illegible] [illegible] каждой выборки будет, по крайней мере [illegible] средних других выборок. С другой стороны [illegible] что выборки извлекались согласно МРВО, не каждая [illegible] репрезентативной для данной генеральной [illegible] будем продолжать извлекать выборки объемом 100, что [illegible] попадется выборка, состоящая только [illegible] молодых жителей. С[illegible] этой выборки б[illegible] [illegible] аналогично некоторые из на[illegible] выбо[illegible] [illegible] людей и иметь сре[illegible] [illegible] здравый смысл говорит [illegible] [illegible] ином сред[illegible]

[illegible] таким образом узнали, что истинное сред[illegible] составляет 30 лет. Как показано выше, сред[illegible] равны 30. Следовательно, рис[illegible] должно иметь пик на 30 годах. Некоторые [illegible] частота таких промахов должна убывать

[illegible] не выборок, равные 29 или 31, [illegible] встречаться, средние выборок, равные 20 или 40, должны встречаться [illegible] выборки случайные, то их средние должны промахиваться одинаково [illegible] стороне от истинного среднего генеральной совокупности [illegible], само распределение должно быть примерно симметричным [illegible] выборочных средних должно быть приблизительно [illegible] им и напоминать распределение, показанное на рис. 6.3.

Рис. 6.3
[illegible]



[illegible] просе сформулированы в виде двух теорем. Первая теорема [illegible]

Если из генеральной со[illegible], имеющей нормальное распределение со средним μ и стандартным отклонением σ, [illegible] выборки объема N, то распределение вы[illegible] будет нормальным со средним μ и стандартным отклонением $\sigma/\sqrt{N}$.

Поясним теорему: если мы возьмем признак, который в генеральной совокупности подчиняется нормальному распределению (например, IQ, рост, вес), и извлечем для него из этой генеральной совокупности бесконечное количество случайных выборок одинакового объема, то распределение полученных нами выборочных средних будет нормальным. Если мы знаем, что какая-то переменная распределена нормально [illegible] что и распределение всех возможных выборочных средних является нормальным.

Данная теорема [illegible] нам не только форму распределения выборочных средних, но также и значения его среднего и стандартного отклонения. Фактически теорема [illegible] распределения будет в точности таким же [illegible] летнее [illegible] генеральной совокупности [illegible] пределения выборочных средних 10 также будет равно 100. Почему это так — мы [illegible] [illegible] выборочных средних [illegible]

Тара-ра, тара-ра.

[illegible] генеральной совокупности, деленному на корень [illegible] из объема выборки [illegible].

[illegible]

такой объем информации (форма, центр распределения, разброс), который [illegible] ю массы для любого эмпирического распределения. Но что нам делать в [illegible] распределение генеральной совокупности не является нормальным? Такой [illegible] распределение генеральной совокупности не является нормальным. Такой случай (фактически очень часто встречающийся) подпадает под условия второй [illegible] центральной предельной теоремой

"Если из генеральной совокупности, имеющей любое распределение [illegible] выборки объема N, то при большом N распределение [illegible] выборочных средних будет стремиться к нормальному распределению со средним μ стандартным отклонением $\sigma/\sqrt{N}$."

Важность центральной предельной теоремы состоит в том, что она снимает требование нормальности распределения генеральной совокупности. Ес[illegible] борки большой, то можно принять допущение о том, что распределение выборочных средних является нормальным со средним, равным среднему генеральной [illegible]

[illegible] имеем право принять допущение о нормальности распределения выборочных средних.

[illegible]

6.5 Обозначения и терминология

[illegible]

[illegible]

[illegible]

генеральных совокупностей. В таблице 6.1 приведены некоторые символы для обозначения распределения выборочных статистик в кратком виде для быстрой справки. В основном распределение выборочных статистик обозначается греческими [illegible] ами с индексом, соответствующим интересующей нас выборочной статистике.

Таблица 6.1. Символы, применяемые для обозначения средних и стандартных отклонений трех различных распределений

	Среднее	Стандартное отклонение	Пропорция
1. Выборки	$\bar{X}$	[illegible]	P_s
2. Генеральные совокупности	μ	σ	P_u
3. Распределения [illegible]			
выборочных средних	$\mu_{\bar{X}}$	$\sigma_{\bar{X}}$	
выборочных пропорций	μ_p	σ_p	

Заключение

1. Так как генеральные совокупности практически всегда слишком велики, чтобы их можно было протестировать целиком, то фундаментальной стратегией исследования [illegible] нас генеральной совокупности, а затем использование информации из этой выборки для формулирования обобщений на всю генеральную совокупность. Это производится путем построения оценок или путем проверки гипотез.
2. Широко распространены несколько разновидностей метода формирования случайных выборок. Каждая из них включает отбор наблюдений в выборку в соответствии с принципом МРВО. Однако даже наиболее строгий метод не может гарантировать репрезентативность выборки. Одним из существенных достоинств статистических выводов является то, что вероятность ошибки такого рода (а именно — не репрезентативности выборки) можно оценить.
3. Выборочное распределение [illegible] статистики — это центральное понятие в статистических выводах. Это распределение является теоретическим. Исследователь может адекватно охарактеризовать и [illegible] распределение всех возможных выборочных методов [illegible] для средних, как известно его [illegible] форма [illegible] (при соблюдении условий, указанных в первой или второй теореме).
4. [illegible]

[illegible] пропорции, будут введены далее по мере [illegible].

[illegible] объем выборки достаточно велик, то тогда [illegible] выборочных средних будет нормальным по форме [illegible] будет равняться стандартному отклонению (стандартная ошибка) будет равняться стандартному отклонению генеральной совокупности, деленному на корень квадратный из N.

5. [illegible] Как [illegible], так и проверка гипотез используют три совершенно различные распределения [illegible]

Глоссарий

Групповая (кластерная) выборка. Это выборка, сформированная методом, при котором случайно отбираются географические единицы, а затем отбираются в выборку все наблюдения внутри каждой отобранной единицы.

Параметр. Некоторая характеристика генеральной совокупности.

Простая случайная выборка. Это выборка, сформированная методом отбора наблюдений из генеральной совокупности, при котором каждое наблюдение и каждая комбинация наблюдений имеют равную вероятность того, что их включат в выборку.

Распределение всех выборочных статистик. При условиях, указанных в двух теоремах, распределение выборочных статистик будет нормальным по форме, его среднее будет равняться среднему генеральной совокупности, а стандартное отклонение (стандартная ошибка) будет равняться стандартному отклонению генеральной совокупности, деленному на корень квадратный из N.

Расслоенная (стратифицированная) выборка. Это выборка, сформированная методом, при котором наблюдения [illegible] из подсписков генеральной совокупности.

Репрезентативность. [illegible] Выборка является [illegible]

Систематическая выборка. Это выборка, сформированная методом, при котором [illegible] списка элементов генеральной совокупности. [illegible] пока через каждые k номеров [illegible]

Стандартная ошибка среднего. Это стандартное отклонение распределения выборочных [illegible]

Центральная предельная теорема. Теорема, [illegible] распределения всех возможных выборочных средних [illegible] объема выборки.

МРВО. Расшифровывается как Equal Probability of SElection Method — "Метод равновероятностного отбора — МРВО". [illegible] генеральной совокупности.

P_s (P с индексом s). Любая пропорция в выборке

P_u (P с индексом u). Любая пропорция в генеральной совокупности

μ Среднее генеральной совокупности

$\mu_{\bar{X}}$ Среднее распределения выборочных средних

μ_p Среднее распределения выборочных пропорций

Мультимедийные ресурсы

The Wadsworth Sociology Resource Center Virtual Society (Виртуальное сообщество Центра социологических ресурсов Вадсворта)

[illegible]

Посетите сайт, сопровождающий шестое издание книги *Statistics: A Tool for Social Research* (*Статистика* [illegible]) [illegible] широкий спектр ресурсов для изучающих статистику. Начните со щелчка на разделе *Student Resources* этого сайта, чтобы получить доступ к таким материалам:

- Обзор основ математики
- Обзор статистики
- Обучающие карточки
- Ссылки на полезные интернет-ресурсы
- [illegible]
- Таблицы [illegible]
- Примеры и упражнения для MicroCase и SPSS
- Функциональные диаграммы "Find the test"
- Проверка гипотез для переменных, измеряемых по порядковой шкале

Задачи

6.1. Пусть вам нужно сформировать случайную выборку ($N = 300$) из студентов [illegible] [illegible] выборки вы примените (т.е. сформируете простую выборку [illegible]) [illegible] Является ли групповая (кластерная) выборка подходящей [illegible] формирование групповой выборки.

6.2. [illegible] глубокое и отчетливое понимание того, что [illegible], если вы хотите получить более глубокое и отчетливое понимание того, что [illegible] выборочных статистик и две теоремы, [illegible] случайный отбор (например, с помощью таблицы [illegible])

[illegible] крайней мере, 50 выборок объемом 2 элемента (т.е. 50 пар наблюдений) [illegible] рассчитайте среднее каждой выборки. Постройте по этим средним [illegible] полигон частот (ломаную линию, получаемую при соединении точек, лежащих по [illegible] классов [illegible] более показательным [illegible] выборки, больший чем $N=2$.

а. Построенная вами кривая (ломаная) будет представлять собой распределение всех возможных выборочных средних. Какова форма вашей кривой? Для 50 выборок она уже должна приближаться к нормальному распределению. Исходя из формы вашей кривой, какова будет ваша оценка среднего генеральной совокупности (μ)?

б. Рассчитайте среднее выборочного распределения всех возможных выборочных средних ($\mu_{\bar{Y}}$). Будьте внимательны. Для этого просуммируйте выборочные средние и разделите полученную сумму на количество извлеченных вами выборок. Теперь рассчитайте среднее генеральной совокупности (μ). Эти два средних [illegible] приблизительно одинаковыми [illegible] предельной теореме $\mu=\mu_{\bar{Y}}$

в. Рассчитайте стандартное отклонение для распределения выборочных средних (при этом для расчета используйте средние арифметические выборок). Рассчитайте стандартное отклонение для генеральной совокупности [illegible] σ [illegible] σ. Вы должны обнаружить, что [illegible] выполняется соотношение $\sigma_{\bar{Y}}=\sigma/\sqrt{N}$.

г. [illegible] у вас не получилось, то это [illegible] за одной или нескольких нижеуказанных причин:

1. Вы не извлекли достаточно большого количества выборок. Вам может понадобиться 100, 200 или даже больше выборок, чтобы увидеть, что кривая начинает принимать вид кривой нормального распределения.
2. Размер выборки ($N=2$) слишком мал. Попробуйте $N=5$ [illegible]
3. Ваш метод формирования выборки не является случайным, и/или элементы вашей генеральной совокупности [illegible] случайным образом.

7	21	20	19	21
14	7	7	21	19
19	22	19	23	19
21	23	18	20	20
[illegible]	[illegible]	[illegible]	[illegible]	20
[illegible]	[illegible]	8	2	[illegible]
[illegible]	18	21	9	21
22	7	21	[illegible]	2
3	[illegible]	[illegible]	20	[illegible]
8	2	20	3	21

SPSS for Windows

Использование *SPSS for Windows* для формирования случайных выборок

Демонстрационный пример с использованием SPSS 6.1. Оценка среднего возраста

В *SPSS for Windows* есть процедура для извлечения случайных выборок из файла данных. Мы воспользуемся этой процедурой для иллюстрации некоторых моментов, касающихся формирования выборок. [illegible] рами генеральной совокупности.

Ниже приведены инструкции, следуя которым вы рассчитаете средние для переменной *age* [illegible] из выборки GSS 1998 г. Истинный средний возраст для выборки GSS 1998 г. составляет $\mu=44,94$ (см. демонстрационный пример 5.1). Объемы наших трех выборок приблизительно составляют 10, 25 и 50% от объема генеральной совокупности. SPSS сформирует эти выборки с помощью процесса, похожего на таблицу случайных чисел. Следовательно, эти выборки можно считать "простыми случайными выборками".

Как часть процедуры, мы также потребуем рассчитать "стандартную ошибку среднего" (S.E. mean). Это стандартное отклонение для распределения выборочных средних $\sigma/\sqrt{N}$ для выборок заданного размера. Нас интересует эта статистика, поскольку [illegible] ности не далее, чем на величину стандартной ошибки среднего.

Загрузите файл данных GSS 1998 г. Выберите из главного меню **Data** (Данные) и [illegible] Select Cases [illegible] борку, поэтому пометьте элемент **Random sample of cases** (Случайная выборка) [illegible] Sample [illegible] Select Cases: Random Sample [illegible] двух [illegible] способ: указать объем выборки. Давайте воспользуемся пер[illegible] буем, чтобы выборка содержала 10% всех наблюдений файла [illegible] лем 10 в поле ввода первой строки нашего окна. [illegible] OK [illegible] должить) окна Select Cases: Random Sample [illegible] Теперь выборка сформирована [illegible] Analyze [illegible]

Чтобы найти среднее нашей [illegible] Descriptive Statistics [illegible] Descriptives [illegible]

[illegible] Variable(s) справа. Щелкните по кнопке Options [illegible] Continue и затем OK, после чего [illegible] описательными статистиками.

Теперь, чтобы сформировать 25%-ную выборку, возвратитесь в окно Select [illegible] Reset [illegible] OK [illegible] random sample of cases [illegible] Sample [illegible] за данных, для этого введите 25 в поле ввода первой строки [illegible] Continue и затем OK, и сформируется новая выборка.

[illegible] Descriptives для 25%-ной выборки [illegible] тистику S.E.mean. Обратите внимание на окно результатов. Наконец, повторите все [illegible] для 50%-ной выборки. Полученные результаты для трех выборок [illegible] таблица.

Выборка	Объем выборки	Среднее выборки	Стандартная ошибка среднего S.E.mean	Среднее выборки плюс/минус стандартная ошибка среднего
[illegible]%	142	43,65	1,41	42,24–45,06
25%	318	44,42	0,95	43,47–45,37
50%	685	44,93	0,65	44,28–45,58

Отметим, что стандартная ошибка среднего (т.е. стандартное отклонение для распределения выборочных средних) уменьшается при увеличении объема выборки. Здравый смысл говорит нам о том, что большие выборки позволяют получить более [illegible] ние к значению этого параметра ге[illegible] Для каждой из трех выборок [illegible] покупности на расстояние [illegible]

Более того, выборка [illegible] 1,29 года меньше, чем истинное значение [illegible] %-ная выборка позволила получить более [illegible]

Упражнения

6.1. Следуя процедурам демонстрационного примера 6.1, сформируйте три выборки [illegible] базы данных GSS 1998 г. объемом 15%, 30% и 60% от объема генеральной совокупности соответственно. Рассчитайте описательные статистики для [illegible] стандартную ошибку среднего (S.E.mean). Внесите полученные вами значения в следующую таблицу.

Выборка	Объем выборки	Среднее выборки	Стандартная ошибка среднего S.E.mean	Среднее выборки плюс/минус стандартная ошибка среднего
[illegible]				
[illegible]				
[illegible]				

[illegible] средние выборок? Удалены ли все три выборочных средних на расстояние, не превышающее [illegible] стандартной ошибки среднего, от среднего генеральной совокупности (44,94)? Как меняется точность оценок при изменении объема выборки?

7 Процедуры оценивания

7.1 Введение

7.2. Смещение и эффективность

[illegible] ых статистик — распределение всех возможных значений [illegible] выборочных средних. Таким образом, очень хо- [illegible]

[illegible] что оно находится в пределах ±3 Z от среднего значения рас- [illegible]

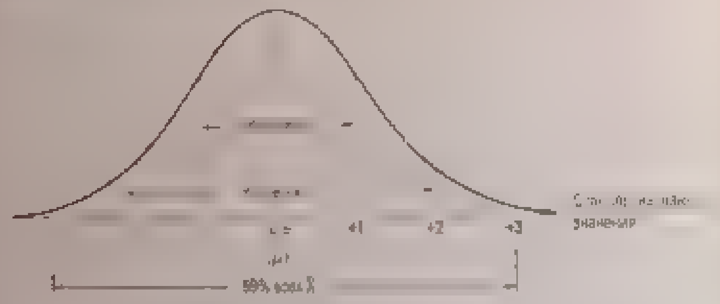


Рис. 7.1. [illegible] под распределением средних значений выборок

Если статистическая оценка является несмещенной, она, вероятно, представляет собой точную оценку параметра генеральной совокупности (в данном случае, μ) [illegible] внимание на то, что менее чем для 1% случаев выборочное среднее [illegible] более чем ±3 Z от среднего значения распределения выбороч- [illegible] Мы не имеем ни малейшего понятия, попадает наше выборочное сред- [illegible]

[illegible] характеристикой статистической оценки является эффективн- [illegible] группирования выборочного распределения около его [illegible] или степень группирования по существу связана с разбро- [illegible] отклонение распределения выборочных статистик, тем [illegible] и выше эффективность. Помните, что стандартное [illegible] отклонению для генеральной совокупности, разделенному [illegible] $= \sigma/\sqrt{N}$). Таким образом, стандартное отклонение [illegible] пропорционально N. При увеличении [illegible] Мы можем увеличить эффективность (т.е. [illegible] выборочных средних) любой ста- [illegible]

Это можно лучше понять на примере. Рассмотрим две выборки различного раз- [illegible]

[illegible]

В обеих выборках средние значения являются несмещенными, но у какой из этих статистических оценок больше эффективность? Рассмотрим выборку 1 и предположим, для примера, что стандартное отклонение для генеральной совокупности (σ) равно $500. В данном случае стандартное отклонение распределения всех возможных средних значений выборок, где N равно 100, будет равно $\sigma/\sqrt{N}$, или $500/\sqrt{100}$, или $50. Для выборки 2 стандартное отклонение всех возможных средних значений выборок, где N равно 1000, будет намного меньше. В частности, оно будет равно $500/\sqrt{1000}$, или $15,8.

Распределение выборочных средних 2 является более сгруппированным, чем распределение выборочных средних 1. В действительности, распределение 2 содержит 68% всех возможных средних значений выборок в интервале $\mu \pm 15,8$, тогда как распределению 1 для этого требуется гораздо более широкий интервал $\mu \pm 50,00$. Вероятность того, что оценка, основанная на выборке, состоящей из 1000 наблюдений, будет близкой по значению к параметру генеральной совокупности, гораздо выше, чем соответствующая вероятность для оценки, основанной на выборке, состоящей из 100 наблюдений. На рис. 7.2 и 7.3 эти соотношения представлены графически.

Подводя итоги, можно сказать, что так как стандартное отклонение распределения выборочных статистик является обратной функцией N, то чем больше выборка, тем больше степень группирования и тем выше эффективность. Отсюди можно сказать, что эта связь между размером выборки и стандартным отклонением (меньше- [illegible] условии, что обе были сформированы случайным образом).

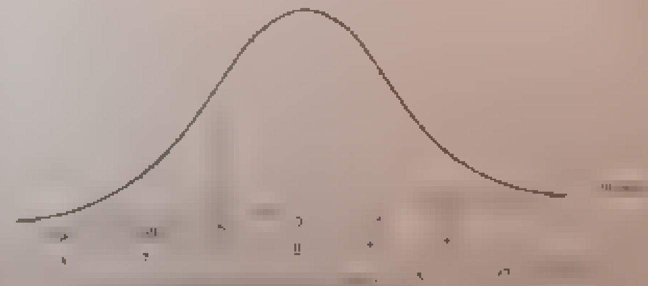


Рис. 7.2 [illegible]

[illegible]

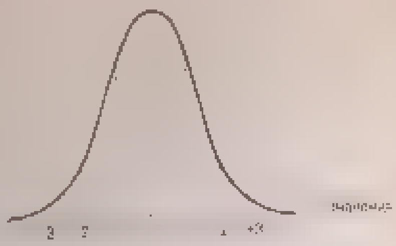


Рис. 7.3
[illegible] средних с N = 1000 и $\sigma_{\bar{x}}$ = $15,81

7.3. Процедуры оценивания: введение

[illegible] оценки довольно проста. В соответствии с [illegible] [illegible] и считаем, что параметр генеральной совокупности равен выборочной статистике. Помните, что чем больше выборка, тем выше [illegible] вероятность того, что статистическая оценка будет приблизительно ра[illegible] для генеральной совокупности. Также помните, что какой бы строгой не была процедура формирования выборки и какой бы большой не была выборка, всегда существует большая вероятность того, что статистическая оценка является неточной.

В сравнении с точечными оценками интервальные оценки более сложны, но и более надежны, потому что когда мы приводим диапазон значений, высока вероятность того, что мы включим в него и параметр, отвечающий всей генеральной совокупности. Первый шаг при построении интервальной оценки заключается в том, чтобы выбрать риск ошибиться, который мы согласны принять. Интервальная оценка является неверной, если она не включает параметр, отвечающий всей генеральной совокупности. Вероятность ошибки называется альфа (обозначается α). Точное значение альфа зависит от природы исследования, но наиболее распространена вероятность, равная 0,05. Если альфа равна 0,05, что соответствует доверительному уровню 95%, это означает, что в перспективе исследовате[illegible] [illegible] конечное число интервалов, то 95% из них будут включать [illegible] ной совокупности, а 5% не будут. В [illegible] [illegible] ем шансы в свою пользу, т.е. в пользу [illegible], что интервал будет включать значен[illegible] [illegible]

Вторым шагом является изображение распределения выборочных статистик, разделение вероятности ошибки на равные правый и левый хвосты кривой распределения и последующее нахождение соответствующих [illegible] например, если мы зададим альфа равной 0,05, тогда мы должны [illegible] (0,025) этой вероятности [illegible] хвосту распределения. Таким образом [illegible] выборочных статистик будет разделено, как показано на рис. 7.4.

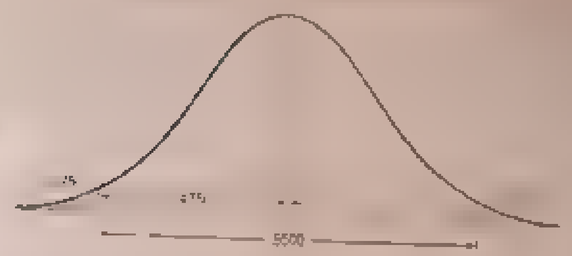


Рис. 7.4
[illegible]

Нужно найти такое стандартизованное значение, которое соответствует краям затененной области. В главе 5 мы изучали, как вычислять стандартизованные значения и как находить площадь под нормальной кривой. Здесь мы применим обратный процесс. Нам нужно найти стандартизованное значение, за пределами которого лежит 0,0250 общей площади. Чтобы это сделать, будем просматривать столбец C приложения A, пока не найдем это значение (0,0250). Соответствующее ему стандартизованное значение равно 1,96. Так как кривая симметрична, и нас интересуют оба хвоста кривой распределения, и левый, и правый, выберем такое стандартизованное значение, которое соответствовало бы альфе, равной 0,05, т.е. ±1,96 (см. рис. 7.5).

Рис. 7.5 [illegible]

[illegible] мы можем быть на 95% уверенными в том [illegible] [illegible] уровня [illegible] доверительных уровней: 90% (α = 0,10), 99% (α = 0,01) и [illegible] [illegible] де 7.1 приведена вся нужная нам информация.

Таблица 7.1. Стандартизованные значения для различных уровней альфа (α)

Доверительный уровень	Альфа	α/2	Стандартизованное значение
90%	0,10	0,0500	±1,65
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
99,9%	0,001	0,0005	±3,29

Вам следует обратиться к приложению А и проверить, действительно ли стандартизованные значения из таблицы 7.1 соответствуют этим уровням альфа. При этом обратите внимание на то, что у наблюдений, для которых альфа равна 0. 0 и 0.01 точных значений искомой нами площади в таблице нет. Например, если альфа равна 0.10, мы можем обратиться к столбцу С ("Площадь за пределами"), надеясь найти там значение 0.0500. Вместо этого мы видим там значение площади 0.0505 ($Z = \pm 1,64$) и значение площади 0.0495 ($Z = \pm ,65$). Искомое нами стандартизованное [illegible] находится где-то между этими двумя значениями. Когда возникает такая ситуация, следует брать большее из двух стандартизованных значений. Это сделает интервал шире при данных обстоятельствах, а потому такой вариант является более осторожным. В случаях, когда альфа равна 0.01, мы сталкиваемся с такой же проблемой (точного значения площади 0,0050 в таблице нет), поэтому нам следует решить ее подобным образом и выбрать большее из значений Z. В конечном итоге, обратите внимание на то, что в случае, когда альфа равна 0.001, мы можем выбрать из нескольких стандартизованных значений. Хотя наша таблица не является достаточно подробной, чтобы это показать, ближайшее стандартизованное значение соответствует [illegible] до ±3,29. *В качестве практического задания, касающегося нахождения стандартизованных [illegible]* [illegible]

7.4. Процедуры интервального оценивания выборочных средних

Формула построения доверительного интервала, основанного на выборочных средних, имеет следующий вид:

Формула 7.1

$$\text{д.и.} = \bar{X} \pm Z \frac{\sigma}{\sqrt{N}}$$

[illegible]

Z — [illegible], определяемое уровнем альфа

[illegible] распределения выборочных средних

[illegible]

Для примера, предположим, что вы хотите [illegible] коэффициента развития интеллекта [illegible] IQ [illegible] средние значения [illegible] стандартное отклонение значений IQ равно [illegible] $\sigma = 15$ [illegible] стандартизованное значение будет равно 1,96. Подставив эти [illegible] ву 7.1 [illegible]

$$\text{д.и.} = \bar{X} \pm Z \frac{\sigma}{\sqrt{N}}$$

[illegible]

$$\text{д.и.} = 105 \pm (1,96)(1,06)$$

$$\text{д.и.} = 105 \pm 2,08$$

Таким образом, наша оценка показывает, что истинное среднее значение коэффициента IQ для рассматриваемой генеральной совокупности находится где-то между 102,92 (105 – 2,08) и 107,08 (105 + 2,08). Так как 95% всех возможных выборочных средних находится в пределах ±1,96 Z (в данном случае, 2,08 единиц IQ) от среднего значения распределения выборочных средних, велики шансы того, что наш интервал будет включать среднее значение генеральной совокупности. В действительности, даже если выборочное среднее будет находиться на расстоянии ±1,96 Z (что маловероятно), наш интервал все еще будет содержать $\mu_{\bar{X}}$, а значит и μ. Нам не удастся включить среднее значение генеральной совокупности только в том случае, если наше выборочное среднее будет одним из тех немногих значений, которые находятся за пределами ±1,96 Z от среднего значения распределения выборочных средних.

Обратите внимание на то, что в примере, приведенном выше, у нас было значение [illegible] что то, что мы обладаем такой информацией о генеральной совокупности, довольно необычно. В абсолютном большинстве случаев мы не будем знать значения σ [illegible] стандартным отклонением). К сожалению, [illegible] размеры смещения [illegible] ванная формула для случаев с неизвестным значением σ имеет [illegible]

Формула 7.2 [illegible]

[illegible] ... стандартное нормальное распределение, приведенное в приложе[illegible] ... Существует возмож[illegible] ... [illegible] меньше 100 наблюдений, но для этого нам следует использовать другое теоретическое распределение, которое называется t-распределением Стьюдента, для нахождения площади под распределением выборочных статистик. Мы отложим рассмотрение t-распределения до главы 8, а пока заострим свое внимание на процедурах оценивания, предназначенных только для больших выборок.

В заключение этого раздела давайте рассмотрим задачу-пример, в которой используется формула 7.2. Средний доход в случайной выборке из некоторой группы людей составляет $35000, причем стандартное отклонение равно $200. Чему равна 95%-ная интервальная оценка среднего значения генеральной совокупности, μ?

Исходя из того, что:

$\bar{X} = \$35000$

[illegible]

$N = 500$

[illegible]

$$\mu = \bar{X} \pm Z\left(\frac{s}{\sqrt{N-1}}\right)$$

$$= [illegible] \left(\frac{200}{\sqrt{499}}\right)$$

$$[illegible]$$

[illegible] для этой группы находится в пределах от $34982.45 [illegible]

7.5. Процедуры интервального оценивания для выборочных пропорций (большие выборки)

[illegible] ся в том, что так как пропорции представляют собой статистический показатель, [illegible] выборочных статистик. В действительности, опять же исходя из центральной предельной теоремы, мы знаем, что пропорции выборки имеют нормальные распределения [illegible] Формула построения доверительных интервалов на основании выборочных пропорций имеет следующий вид:

Формула 7.3 [illegible]

Значения P_s и N мы получаем напрямую из выборки, а значение Z определяется доверительным уровнем, как и в случае со средними значениями выборок. У нас в формуле остается одна неизвестная, P_u — то значение, которое мы и пытаемся найти. Эту дилемму можно решить, установив для P_u значение 0,5. Так как в числителе второе выражение, находящееся под знаком корня, $(1 - P_u)$, обратно P_u, [illegible] не будет иметь значение 0,5 × 0,5, или 0,25, что представляет собой [illegible] значение, которое только может принять это выражение. То есть, если мы установим для P_u любое значение, отличное от 0,5, значение выражения $P_u(1 - P_u)$ будет уменьшаться. Например, если мы установим для P_u [illegible] $(1 - P_u)$ будет равно 0,6, и значение всего выражения снизится до 0,24. Устанавливая для P_u значение 0,5, мы обеспечиваем, что выражение [illegible]

[illegible] предположим, что нам нужно [illegible] в прошедшем семестре, по меньшей мере, один день занятий [illegible] доверительном уровне 95% интервальная оценка будет такой:

[illegible]

$$\text{д.и.} = 0{,}30 \pm 1{,}96 \sqrt{[illegible]}$$

... равного 0,30, мы можем ... по меньшей мере, один ... 0,23 и 0,37. Оценку, конечно же, можно ... 23% по 37% студентов более ... в течение ...

Практический пример 7.1

... время препровождения после работы ... состоящей из 1000 семей. В качестве ... респонденты указали просмотр ... его значения генеральной совокупности ... продолжительность просмотра теле...

На основании этого результата мы можем вычислить, что все субъекты генераль...

...няя граница (6,2 + 0,04) равна 6,24. Таким образом, этот интервал можно было указать другим способом

Среднее значение генеральной совокупности больше или равно 6,16 и меньше или равно 6,24. Вероятность того, что эта оценка ошибочна (т.е., что она не включает среднее значение генеральной совокупности), равна 5%.

Вы, скорее всего, могли встречать оценки значений генеральной совокупности в средствах массовой информации в форме опросов общественного мнения, прогнозов результатов выборов и т.п. Компании, занимающиеся опросами общественного мнения, используют интервальные оценки, а в средствах массовой информации сообщается сама оценка (например, "В одном из последних опросов американской общественности 45% респондентов сказали, что они одобряют работу президента"), но также указывается ширина интервала ("Точность этой оценки равна ±3%" или "Цифры, полученные в данном опросе, могут содержать ошибку выборки ±3%"), уровень альфа (доверительный уровень обычно равен 95%) и размер выборки ("В опросе приняло участие 458 ...").

В последние десятилетия стали широко применяться прогнозы результатов выборов и анализ намерений избирателей, по крайней мере, на уровне выборов президента. С недавних пор опросы общественного мнения все больше стали использоваться для измерения реакций на все жизненные события, от выхода новых художественных фильмов до сплетен относительно поведения президента во время последнего кризиса ... результаты опросов в качестве дополне... ...знаком телеви... ...их примеров, но опросы стали настолько распространенными ... примеры и самостоятельно. Просто возьмите ... что вы найдете, по меньшей мере, один опрос. Прочтите эту статью ... генеральную совокупность, шир... ... занятие, и вы удивите своего преподавателя.

Как гражданину и как социологу, вам следует очень ... которые не включают такую необходимую инф... интервала. Также вам следует посмотреть, как ... выборка была сформирована неслучайным обр... опрос только среди своих подписчиков) ...

[illegible] эти глубинные выборки [illegible] заинтересованными [illegible] показали или производящие впечатление [illegible] Однако часто такие оценки бывают [illegible] выборок и искаженных процедурах [illegible] образом, которые побуждают [illegible] выборки [illegible] "людьми с улицы" (или покупателями в магазине [illegible]) [illegible] к ним не следует относиться [illegible] с респектабельными исследовательскими [illegible]

Опросы общественного мнения в профессиональной литературе

[illegible] наиболее важное следствие [illegible] в том, что теперь для исследователей [illegible] общенациональные базы данных. Возможность доступа к [illegible] базам данных [illegible] провести высококачественн[illegible] затрат на самостоятельный сбор данных. [illegible] мы можем легко проверять [illegible] включения [illegible] образом [illegible] затраченные ими усилия на проведение [illegible] в обществе [illegible]

[illegible] баз данных такого рода [illegible] исследование General Social Survey (GSS). [illegible] каждый год, начиная с 1972, Национальный совет исследований общественного мнения (National Opinion Research [illegible]) проводит опрос в общенациональной выборке [illegible], в которой [illegible] широкий спектр вопросов. Так как многие [illegible] GSS представляет последовательную запись [illegible] каждый год добавляются [illegible] таким образом множество доступной информации [illegible] общенациональные выборки [illegible] размер выборки варьируется от 1400 до более 3000, и [illegible] основанные на таких больших выборках, будут иметь точность в пределах ±3 [illegible] таблицу 7.4 и [illegible] упражнения, приведенные в этой книге, [illegible] обследовании 1998 года, и эта база [illegible] в приложении Б.

7.6. Управление шириной интервальных оценок

[illegible]

[illegible] больше шансов включить значение для генеральной совокупности. Поэтому мы можем [illegible] доверять.

Чтобы [illegible] 500 [illegible] 53 [illegible] в $200. Мы построили 95%-ный доверительный интервал и обнаружили, что он имеет следующий вид: $35000 ± 17,55.

[illegible] доверительный уровень), стандартизованное значение в формуле уменьшилось [illegible] 1,65, и интервал был бы [illegible]

$$\text{д.и.} = 35{,}000 \pm 1{,}65\left(\frac{200}{\sqrt{499}}\right)$$

$$\text{д.и.} = 35{,}000 \pm ([illegible])([illegible])$$

$$\text{д.и.} = 35{,}000 \pm 14{,}77$$

С другой стороны, если бы мы построили 99%-ный доверительный интервал, [illegible] формуле [illegible] 2,58 [illegible] интервал был бы [illegible]

$$\text{д.и.} = 35{,}000 \pm 2{,}58\left(\frac{200}{\sqrt{499}}\right)$$

$$\text{д.и.} = 35{,}000 \pm (2{,}58)(8{,}95)$$

$$\text{д.и.} = 35{,}000 \pm 23{,}09$$

Таблица 7.2. Интервальные оценки для четырех доверительных уровней ($\bar{X}$ = \$35000, s = \$200, N = 500)

Альфа	Доверительный уровень	Интервал	Ширина интервала
[illegible]	90%	[illegible]	[illegible]
[illegible]	95%	[illegible]	[illegible]
[illegible]	[illegible]	[illegible] 23,09	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

... выборочные средние. Точно такая же связь ... [illegible]

(... и ширине интервалов, см. задачу 7.13.)

Таблица 7.3. Интервальные оценки для четырех различных выборок ($\bar{X}$ = $35000, s = $200, альфа = 0,05)

Выборка 1 (N = 100)	*Выборка 2 (N = 500)*
д.и. = $35000 ± 1,96(200/$\sqrt{99}$)	д.и. = $35000 ± 1,96(200/$\sqrt{499}$)
д.и. = $35000 ± 39,40	д.и. = $35000 ± 17,55
Выборка 3 (N = 1000)	*Выборка 4 (N = 10000)*
[illegible]	[illegible]
д.и. = $35000 ± 12,40	д.и. = $35000 ± 3,92

Выборка	N	Ширина интервала
1	100	$78,80
2	500	$35,10
3	1000	$24,80
4	10000	$7,84

... находится в обратной зависимости от ширины интервала. По ... [illegible] ... приведем пример. В таблице 7 ... [illegible] ... По выборкам получены те же данные ... [illegible] ... установлен доверительный уровень 95%. Отношение ... [illegible] ... и для выборочных пропорций. *Для дополнительной ... ширины интервала, см. задачу 7.14.* ... [illegible] ... линейной зависимости ... [illegible]

... прямую зависит от размера выборки, эта связь указывает на уменьшение ... [illegible] ... полученные на выборке большего размера, не будут намного точнее.

Практический пример 7.2

Если 45% случайной выборки, состоящей из 1000 американцев, говорят, что для них ходьба является основным видом физической деятельности, чему равна оценка значения для генеральной совокупности? Известна следующая информация о [illegible]

$$P_s = 0{,}45$$

$$N = 1000$$

Обратите внимание на то, что процентный показатель занимающихся ходьбой был указан в виде пропорции. Если мы возьмем уровень альфа 0,05, соответствующее стандартизованное значение будет равно ±1,96, и интервальная оценка составит:

$$д.и. = P_s \pm Z\sqrt{\frac{P_u(1-P_u)}{N}}$$

[illegible]

$$д.и. = 0{,}45 \pm 1{,}96\sqrt{0{,}00025}$$

[illegible]

$$д.и. = 0{,}45 \pm 0{,}03$$

Теперь можно считать, что пропорция для генеральной совокупности, для которой ходьба является основным видом физических упражнений, находится ... [illegible] ... является основным видом физических упражнений. Вероятность ... [illegible] ... тервал не включает значение генеральной совокупности, равна ... [illegible]

[illegible]

[illegible] этой информации, ее можно использовать. Помимо прочих возможностей применения [illegible] для отслеживания национальных тенденций и изменений общественного мнения во времени (по крайней мере, начиная с 1972 года — первого года, когда проводился опрос GSS). В недавно опубликованной статье профессор Марк Уорр (Mark Warr) использовал [illegible] что американцы "устали от [illegible] Хотя эти высказывания часто появляются в [illegible] отражают уровень [illegible] в [illegible] общества?

В опросах Gallup и GSS, начиная с середины 1960-х годов, присутствовал пункт, касающийся страха стать жертвой преступления: "Есть ли недалеко отсюда, т.е. в пределах мили, такое место, где бы вы побоялись проходить ночью?" Профессор Уорр построил [illegible] с помощью которой можно проследить изменения [illegible]

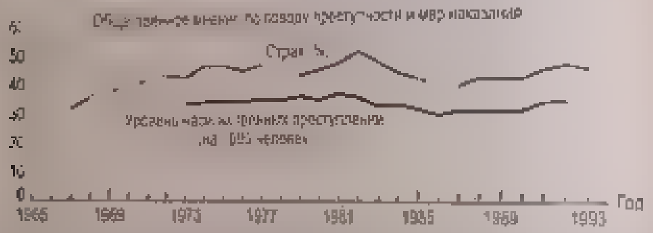


Рис. 1. [illegible] Crime Survey), 1973–92

[illegible] были собраны в случайных выборках [illegible] респондентов [illegible] выводы [illegible] профессор Уорр

Наиболее поражающей чертой [illegible] вопросы исследования [illegible] является относительно [illegible] на протяжении последних двух десятилетий. После небольшого роста в 1960 [illegible] оставался приблизительно на уровне [illegible] в пределах 40 [illegible] стабильность находится в контрасте [illegible] сообщениями [illegible] в средствах массовой информации [illegible]

Профессор Уорр [illegible] общественного мнения и [illegible] к выводу, что хотя американская общественность склонна одобрять [illegible] мер [illegible] мнение является более [illegible]

[illegible] Public Opinion Quarterly [illegible] Перепечатано с разрешения автора и [illegible]

7.7. Интерпретация статистик: Доверительные интервалы, опросы общественного мнения и президентские выборы 2000 года

[illegible] и до дня, когда избиратели пошли в кабину для голосования, исследовательские компании постоянно говорили о том, что кандидаты имеют практически равные шансы, и о том, что имеет место "статистическая ничья". В этой главе, после [illegible] интерпретации статистик, мы исследуем эти утверждения с помощью только что полученных нами знаний о доверительных интервалах.

Таблица 7.4. Результаты опроса Gallup, февраль — ноябрь 2000 года

Дата опроса		Размер выборки (N)	Процент выборки за	
Начало	Конец		Гора	Буша
25 февраля	27 февраля	1004*	43	52
[illegible] марта	12 марта	1006*	43	49
30 марта	2 апреля	996*	45	46
7 апреля	9 апреля	502	41	50
28 апреля	30 апреля	499	44	48
18 мая	21 мая	1011*	38	45
6 июня	7 июня	528	44	48
23 июня	25 июня	1020*	38	52
14 июля	16 июля	535	40	48
[illegible]	5 августа	650	37	54
18 августа	19 августа	1043*	47	46
8 сентября	10 сентября	727	49	42
2[illegible] сентября	23 сентября	693	46	47
28 сентября	30 сентября	691	45	45
20 октября	22 октября	769	44	46
30 октября	1 ноября	2123	43	47
5 ноября	6 ноября	2350	46	48

*[illegible]

Данные

[illegible] и множество различных агентств [illegible]

Полный список [illegible] представлен на сайте www.gallup.com/poll/releases.

[illegible]

с каком политическом [illegible]

[illegible] опросов.

В таблице 7.4 представлены результаты 17 опросов, проведенных Gallup с февра- [illegible]

[illegible]

кандидатуры от соответствующих партий. С февраля до начала [illegible]

[illegible] Вторая фаза началась в конце лета после Республиканской и Демократической нацио- [illegible]

середины сентября, преимущество перешло к Гору. Во время последнего этапа, с середины сентября до дня выборов, Буш вновь занял лидирующее, однако в очень незначительной степени, положение. Как мы видели в главе 2, графики являются более эффективным средством сообщения результатов, чем таблицы. На рис. 7.6 представлены данные из таблицы 7.4 и видны все три этапа соперничества между [illegible]

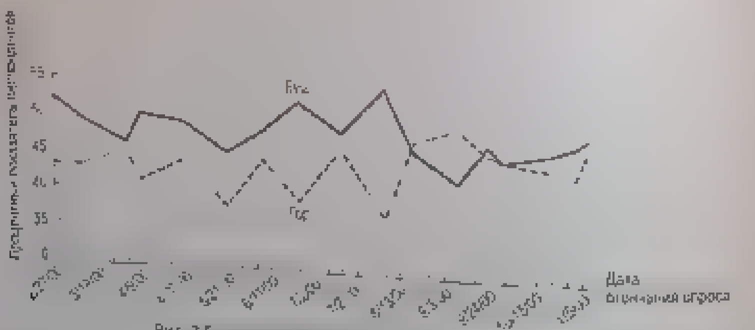


Рис. 7.6. [illegible]

Доверительные интервалы

[illegible]

[illegible]

интервалы для каждого кандидата, и, наконец, ширина интервальных оценок. На- [illegible] голосовать за Гора. Заменив процентные показатели пропорциями, и подставив в формулу 7.3 соответствующие значения, мы получим доверительный интервал сле- [illegible]

[illegible]

0,43 ± 0,031

[illegible] обратно в проценты, интервал будет иметь следующий вид

[illegible]

Этот интервал простирается от 39,9 до 46,1%, а его ширина равна 6,2%.

Таблица 7.5 подтверждает очень важную мысль относительно оценивания: при увеличении размера выборки сужается ширина интервала (см. также таблицу 7.3). Например, некоторые опросы были основаны на сравнительно небольших выборках, состоящих из приблизительно 500–600 вероятных избирателей. В таких случаях интервалы имели ширину 8,8%. В контрасте с этим последние два опроса, проводившиеся в ноябре, были основаны на выборках гораздо большего размера [illegible] как следствие, интервалы имели ширину 4%, т.е. более чем в два раза уже, чем для вы- [illegible] борки почти в четыре раза больше, чем выборки меньшего размера, соответствую- [illegible] лиц 7.6).

Таблица 7.5. Размер выборки, доверительные интервалы и вероятные победители

Дата окончания опроса	Размер выборки	95%-ный доверительный интервал за Гора	95%-ный доверительный интервал за Буша	Ширина интервала
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	45 [illegible]	[illegible]	[illegible]
[illegible]	502	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

Дата окончания опроса	Размер выборки	95%-ный доверительный интервал за Гора	95%-ный доверительный интервал за Буша	Ширина интервала
[illegible]	[illegible]	[illegible]	45 [illegible]	2
[illegible]	[illegible]	[illegible]	46 [illegible]	8.6
[illegible]	[illegible]	[illegible]	[illegible]	6.2
[illegible]	[illegible]	[illegible]	[illegible]	8
[illegible]	[illegible]	[illegible]	[illegible] 3.8	7.6
[illegible]	[illegible]	[illegible]	[illegible] 3.0	6.0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	4
[illegible]	[illegible]	[illegible]	[illegible]	7.4
[illegible]	[illegible]	[illegible]	[illegible]	7.0
[illegible]	[illegible]	[illegible]	[illegible]	4.2
[illegible]	[illegible]	[illegible]	[illegible]	6

[illegible]

Самой важной информацией, содержащейся в таблице 7.5, являются интервалы сами по себе. До дня выборов оставалось все меньше и меньше времени, так почему же было столько неясности относительно того, кто занимает лидирующее положение? Почему результаты были настолько близки? Помните, что прогнозы относительно "вероятного победителя" должны быть основаны на сравнении доверительных интервалов (а не точечных оценок), и если интервалы перекрываются, существует возможность, что кандидат, который кажется отстающим, на самом деле [illegible]

Например, самый последний опрос, проведенный в ноябре, показал, что Гора [illegible] Гора поддерживало всего 44% а Буша [illegible] большинство опросов, проводимых с февраля, показали, [illegible]

[illegible] Когда были окончательно подсчитаны [illegible] между избирателями действительно [illegible] их отделяли лишь несколько тысяч [illegible] число равно 100 миллионов. Гор получил [illegible] больше, чем верхний предел оценки на 6 ноября, тогда [illegible] параметр входящий в последнюю ноябрьскую [illegible]

Заключение

1. Значения генеральной совокупности можно оценить, используя значения выборки. [illegible] Доверительный интервал [illegible] что значение генеральной совокупности попадает в некоторый диапазон значений.
2. [illegible] выборки. Чем больше размер выборки, тем [illegible] нения распределения [illegible]
3. Используя точечные оценки, мы говорим, что значение генеральной совокупности равно выборочной статистике (среднему значению или пропорции). Используя интервальные оценки, мы строим доверительный интервал, т.е. диапазон значений, в который, по нашему мнению, должно попадать значение генеральной совокупности. Ширина интервала является функцией риска допустить ошибку (уровня альфа), который мы согласны принять, и размера выборки. Расширение интервала происходит по мере снижения вероятности допустить ошибку и по мере уменьшения размера выборки.

Сводка формул

Доверительный интервал для выборочного среднего при известном значении стандартного отклонения в генеральной совокупности:

$$\text{д.и.} = \bar{X} \pm Z\left(\frac{\sigma}{\sqrt{N}}\right) \qquad (7.1)$$

Доверительный интервал для выборочного среднего при неизвестном значении стандартного отклонения в генеральной совокупности:

$$\text{д.и.} = \bar{X} \pm Z\left(\frac{s}{\sqrt{N-1}}\right) \qquad (7.2)$$

Доверительный интервал для выборочной пропорции при большом размере выборки

$$\text{д.и.} = P_s \pm Z\sqrt{\frac{P_u(1-P_u)}{N}} \qquad (7.3)$$

Словарь

[illegible] ...ачение генеральной совокуп- [illegible]

[illegible] распределения ее выборочных значений равно значению интересую- [illegible] генеральной совокупност [illegible]

[illegible]

Мультимедийные ресурсы

[illegible]

[illegible] к широкому диапазону студенческих ресурсов посетите web-сайт, со- [illegible] ...ические ресурсы (Student Resources) web-сайта книги [illegible]

- [illegible] основ математики,
- [illegible]

Задачи

[illegible]

а. $\bar{X} = 5{,}2$, $s = 0{,}7$, $N = 157$

б. $\bar{X} = 100$, $s = 9$, $N = 620$

в. $\bar{X} = 20$, $s = 3$, $N = 220$

г. $\bar{X} = 1020$, $s = 50$, $N = 329$

д. $\bar{X} = 7{,}3$, $s = 1{,}2$, $N = 105$

е. $\bar{X} = 33$, $s = 6$, $N = 220$

7.2. Для каждого набора значений выборочных статистик, приведенных ниже, по- [illegible] 99%-ный доверительный интервал [illegible]

а. P [illegible], $N = 100$

б. P [illegible], $N = 522$

в. P [illegible], $N = 121$

г. P [illegible], $N = 1049$

д. P [illegible], $N = 548$

е. P [illegible], $N = 300$

7.3. Для каждого из представленных ниже доверительных уровней определи [illegible] соответствующее стандартизованное значение.

Доверительный уровень	Альфа	Площадь за пределами Z	Стандартизованное значение
95	0,05	0,0250	±1,96
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			

7.4. SW

Вы разработали серию вопросов для измерения "усталости" со [illegible] ников. Среднее значение в случайной выборке из 100 социальных ра [illegible] работающих в городе Ши [illegible], Канзас, составило 10,3, со ст [illegible] клонением в 2,7. Чему равна оценка среднего значен [illegible] ной совокупности в целом? Используйте 95%-ный дове [illegible]

7.5. SOC

Исследователь собрал информацию в случайной [illegible] каждой из перечисленных ниже переменных пост [illegible] валы, чтобы оценить [illegible] уровень 90%.

а. В каждой семье живет в сре [illegible] но 0,35

б. Было обнаруже [illegible]

в. Семьи про [illegible]

7.6. SOC

[illegible] телевизионных программ содержала в среднем 2,37 [illegible] на программу. Чему равна [illegible] актов физического насилия на программу с доверительным уровнем 99% для генеральной совокупности с доверительным уровнем 99%?

$$[illegible]$$
$$= [illegible]$$
$$N = 100.$$

7.7. SOC

В случайной выборке из 429 студентов высших учебных заведений был прове[illegible]

а. Они сообщили, что в предыдущем семестре каждый из них потратил в среднем $178,23 на учебники. Выборочное стандартное отклонение для этих данных составляет $15,78. Постройте оценку среднего значения генеральной совокупности при уровне 99%.

б. Они также сообщили, что в предыдущем семестре каждый из них посетил поликлинику в среднем 1,5 раза. Среднеквадратическое отклонение выборки для этих данных составляет 0,3. Постройте оценку среднего значения генеральной совокупности при уровне 99%.

в. В среднем по выборке респонденты пропустили 2,8 дня занятий за семестр [illegible] ляет 1,0. Постройте оценку среднего значения генеральной совокупности при уровне 99%.

г. В среднем по выборке [illegible]

7.8. CJ

Случайная выборка [illegible] города Шинбоун, штат Канзас, показывает [illegible] 50 респондентов стали жертвами насильствен[illegible] 90%-ный доверительный уровень. *Подсказка:* [illegible] формулой 7.3, *вычислите значение выборочной пропорции [illegible] равна частоте, деленной на N*

7.9. SOC

[illegible] 25 из 178 [illegible] 95%.

7.10. PA

[illegible]

7.11. SOC

В случайной выборке [illegible] от 4 до 45% [illegible]

а. 823 респондента согласились с тем, что "Книги и журналы, содержащие открытый секс, ведут к изнасилованиям и прочим сексуальным преступлениям."

б. 6[illegible] респондентов [illegible] по телевизору показывают слишком много открытого секса"

в. 375 респондентов выборки сказали, что за время брака у них хотя бы однажды был секс с кем-то еще, кроме своего супруга

г. 1023 респондента выборки сказали, что у них был секс до первого брака

д. 800 респондентов согласились с тем, что секс до брака — это "всегда непра[illegible]"

7.12. SW

Была сформирована случайная выборка из 100 пациентов, которые участвовали в программе лечения от алкогольной и наркотической зависимости в течение последних 10 лет. Было обнаружено, что 53 пациента, по меньшей мере, однажды, вновь обращались к программе. На уровне 95% постройте оценку пропорции для генеральной совокупности.

7.13. Для данных выборки, представленных ниже, постройте четыре различные интервальные оценки среднего значения генеральной совокупности, по одной для уровней 90, 95, 99 и 99,9%. Что происходит с шириной интервала при увеличении доверительного уровня? Почему?

$$\bar{X} = [illegible]$$
$$s = 10;$$
$$N = 500.$$

7.14. Для каждой из трех выборок, размеры которых указаны ниже, постройте 95%-[illegible] 0,40. Что происходит с шириной интервала при увеличении размера выборки? Почему?

$$P = 0,40$$

Выборка A. $N = 100;$

Выборка B. $N = [illegible]$

Выборка C. $N = [illegible]$

7.15. PS

[illegible]

держивают кандидата А. Можете ли вы предсказать, кто победит? Воспользуя[illegible]

[illegible] Подсказка: какой процент голосов нужен для победы на выбо[illegible] [illegible] совокупности находится в доверительном интервале [illegible] быть где угодно в пределах этого интервала.

$N = 578$

7.16. SOC

Случайной выборке из 1000 зарегистрированных избирателей штата был задан вопрос, поддерживают ли они предложение о наложении запрета на продажу огнестрельного оружия. Если 510 респондентов поддержали это предложение, можно ли с уверенностью 95% сказать, что его поддерживает "большая часть населения"?

7.17. SOC

Студенческие организации и женские объединения Колледжа Святой Алгебры были обеспокоены снижением количества членов, наблюдавшимся в последние несколько лет, и хотят знать, станут ли новые первокурсники плодородной почвой для набора новых членов. Не имея достаточно денег для того, чтобы провести опрос среди всех 1600 первокурсников, они поручили вам [illegible] На уровне [illegible]

7.18. SOC

[illegible] ам потребителей. Ваша работа [illegible] [illegible] у вас вызывает новый экономичный [illegible] производитель говорит, что ему хватает одного [illegible] 78 миль. После проверки значений пробега [illegible] автомобиля, вы обнаружили, что [illegible] галлон при стандартном отклонении [illegible] результаты скорее подтверждают или [illegible]

7.19. SOC

[illegible] приведенного в приложении [illegible] приведенных в этой [illegible]

Для каждой выборочных статистик, из указанных ниже, постройте 95%-ную интервальную оценку значения генеральной совокупности. Обратите внимание на то, что в качестве основы для своих оценок мы будем использовать "выборку из выборки", и нам следует внести поправку на дополнительный источник ошибки. Однако, как вы увидите, выборка все же дает приемлемые оценки [illegible] известно значение параметра генеральной совокупности.

[illegible] приложения Ж.

- **а.** Среднее значение престижа профессии для 1318 респондентов составило 43,87 при стандартном отклонении 13,52.
- **б.** Респонденты ($N = 1334$) сообщили, что они тратят на просмотр телевизора в среднем 2,86 часа в день при стандартном отклонении 2,20.
- **в.** Среднее количество детей для 1385 респондентов составило 1,81 при среднеквадратическом отклонении 1,67.
- **г.** Из 1372 респондентов, ответивших на вопрос, 367 сказали, что они като[illegible]
- **д.** 335 респондентов из 1386 сказали, что они никогда не состояли в браке.
- **е.** Из 830 респондентов, о которых у нас есть информация, 441 сказали, что во время выборов президента в 1996 году они голосовали за Клинтона.
- **ж.** В ответ на вопрос о смертной казни, 925 из 1266 респондентов сказали, что за умышленное убийство следует наказывать смертной казнью.

SPSS for Windows

Построение доверительных интервалов с помощью процедур *SPSS for Windows*

Демонстрационный пример с использованием SPSS 7.1. Обобщение статистической информации для использования при построении доверительных интервалов

SPSS [illegible] хотя некоторые [illegible] Вместо того чтобы использовать эти программы [illegible] ать то, с чем вы уже могли столкнуться [illegible] всегда бывают полезны. Вычисления, [illegible] бы сложны, и вы, вероятно, сможете [illegible] мощью SPSS.

С другой стороны, вряд ли кто [illegible] ление среднего возра[illegible]

кого обследования (GSS) потребует сложения почти 1500 чисел. Масштабы это [illegible]

[illegible]

Descriptives (Описательные статистики) [illegible] Frequencies [illegible]

Descriptives [illegible] Frequencies [illegible] дает частотные распределения для столбца процентных показателей.

Упражнения

7.1 a. SPSS: С помощью команды Descriptives вычислите средние [illegible] дартные отклонения для *papres80*, *age* и [illegible] ных сформулируйте оценки [illegible]

MicroCase: С помощью команды Univariate постройте 95%-ный доверительный интервал для *papres80*, *age* и *income98*.

б. Выразите доверительные интервалы такими словами, как если бы вам нужно было поместить результаты в газетную статью.

7.2 a. SPSS: С помощью команды Frequencies найдите значения выборочных пропорций (преобразуйте процентные показатели в частотные распределения), чтобы они [illegible] ких показателей: пропорция людей с высшим образованием (*degree*), пропорция чернокожих (*race*) и пропорция тех, кто поддерживает контроль оружия (*gunlaw*). Используйте доверительный уровень 95.

MicroCase: [illegible] пропорции [illegible] ния), чтобы оценить параметр генеральной совокупности [illegible] пропорция чернокожих (*race*) и пропорция тех, кто поддерживает контроль оружия (*gunlaw*). Используйте доверительный уровень [illegible]

б. Выразите [illegible] указать результаты в газетной статье.

Примечание: ваши оценки возрастов не будут [illegible] Выборка GSS [illegible]

8 Проверка статистических гипотез I. Случай одной выборки

Цели изучения К концу этой главы вы [illegible]

1. Объяснить логику проверки статистических ги[illegible]
2. [illegible] проверки статистических гипотез, а именно [illegible] распределение выборочных статистик [illegible] и [illegible] ика критерия.
3. Объяснить [illegible]
4. [illegible]
5. [illegible] значимость средних значений и [illegible]
6. [illegible] критериями и [illegible]
7. Определить и объяснить ошибки [illegible] дой из них с выбором уровня ал[illegible]

8.1. Введение

[illegible] ванием другой области применения статистической [illegible] проверка статистических гипотез или проверка значимости [illegible] процедуры можно использовать [illegible]

1. [illegible] Также него есть информация о [illegible]

[illegible] выборке [illegible]

[illegible] преступлений [illegible]

[illegible]

[illegible] нам следует сравнить с генеральной совокупностью (все жители штата, все [illegible] или все жители города). Обратите внимание на то, что нас, по сути, не интересует сама выборка, нас интересует большая группа, из которой эта выборка была сформирована (все пожилые граждане штата, сотрудники колледжа или люди, [illegible]

[illegible] средний балл или количество [illegible]

[illegible] которые проводили [illegible]

[illegible] показали для выборки, состоящей из пожилых граждан, оказался выше, чем [illegible] по штату, мы можем заключить [illegible] более вероятной жертвой престу[illegible]. Это означает, что очень малове[illegible]

[illegible] небольшое различие [illegible] средним баллом студентов в целом [illegible] сотрудники данного колледжа [illegible]

Таким образом, мы можем использовать выборк[illegible] для большей группы (пожилые граждане, со[illegible] характеристики генеральной совокупности [illegible] Однако помните, что процесс формирования [illegible] основанный на принципе МРВО не гарантирует репрезентативности, и поэтому [illegible] всегда будут содержать [illegible]

8.2 Обзор процесса проверки статистических гипотез

Начнем с общего обзора проверки статистических гипотез, используя в качестве примера третью из вышеперечисленных исследовательских ситуаций. В оставшейся части главы мы рассмотрим соображения технического характера и принимую терминологию. Уровни прогулов для выборки клиентов и города в целом представлены [illegible]

[illegible]	Выборка
[illegible] 7,2 [illegible]	$\bar{X}$ = 6,8 [illegible]
$\sigma = 1{,}43$	$N = 127$

Мы можем видеть, что имеет место разница в значениях уровня прогулов, а также среднее значение уровня прогулов для выборки почти на полдня меньше, чем для генеральной совокупности. Можно ли сделать вывод, что для ВСЕХ алкоголиков, прошедших курс лечения, имеет место более низкий уровень прогулов? Так как мы работаем со случайной выборкой (а не со всеми алкоголиками, принимавшими участие в программе), существуют два возможных объяснения различия уровней прогулов.

Первое объяснение, которое мы обозначим как объяснение A, заключается в том, что различие между 7,2 днями и 6,8 днями отражает реально существующее различие [illegible] прошедших курс лечения [illegible] различие является статистически значимым и очень маловероятно, чтобы оно было вызвано только лишь случайностью. Если объяснение A является истинным, [illegible] уровнем прогулов, равным 7,2 дня.

Второе объяснение, или объяснение B, заключается в том, что наблюдаемое различие [illegible] и обусловлено простой случайностью. Если [illegible] и что выборка извлечена из [illegible] уровнем прогулов, равным 7,2 [illegible]

[illegible]

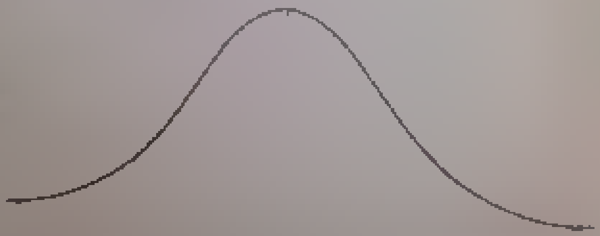
Процесс принятия решения, в первом приближении, начинается с предположе- [illegible]

Помните, что в данном случае μ относится ко всем алкоголикам, прошедшим курс лечения. Это предположение $\mu = 7,2$ можно проверить статистически.

Если объяснение В (алкоголики, прошедшие курс лечения, не отличаются от общества в целом) является истинным, то можно найти вероятность получения наблюдаемого среднего значения выборки ($\bar{X} = 6,8$). Давайте сразу же добавим праг[illegible] чем 0,05 (5 из 100, или 1 из 20) следует отклонить объяснение В. Если это объясне- [illegible] гипотез мы всегда выступаем *против* редких событий.

[illegible] проблемой является определение вероятности наблюдения значения [illegible] статистики, подтверждающей объяснение В. Мы можем определить эту [illegible] с высокой степенью точности, основываясь на том, что мы знаем о рас[illegible] подчиняются все возможные значения выборочных статис- [illegible] равно .43, $\sqrt{127}$, так как $\sigma_{\bar{x}} = \sigma/\sqrt{N}$. Мы также знаем, что [illegible] распределение можно интерпретировать как распределение [illegible] и что указанное выше среднее значение выборки ($\bar{X} = 6,8$) [illegible] возможных значений выборочных средних. Распределение [illegible] его средним значением изображено на рис. 8.1

Рис. 8.1 [illegible]



[illegible]

средние значения выборок. В частности, с помощью стандартизованных значений мы можем графически представить ранее сформулированное прагматическое пра[illegible] меньше 0,05 [illegible] зованное значение, эквивалентное данному значению площади, равно ±1,96. Значения площади и стандартизованные значения приведены на рис. 8.2

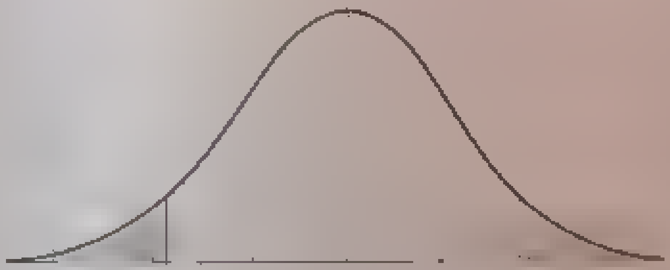


Рис. 8.2 [illegible]

Теперь можно перефразировать прагматическое правило. Вероятность появления [illegible] кривой нормального распределения. Когда мы используем [illegible] ческие распределения, эта формула имеет следующий вид

[illegible]

Чтобы найти стандартизованное значение [illegible] отнимите среднее значение распре[illegible]

Формула 8.1

[illegible]

$$Z = -0,40 \ldots$$

[illegible]

Это Z значение, равное −3,15, указано на изображении распределения всех возможных выборочных средних, приведенном на рис. 8.3, и мы можем наблюдать, что данное значение выборочной статистики попадает в затененную область. Если объяснение B является истинным, то вероятность появления такого значения выборочной статистики меньше, чем 0,05 (то, насколько меньше, не имеет значения, так как решающее правило было установлено заранее). Поэтому, если бы объяснение B было истинным, появление выборочного значения ($\bar{X}$ = 6,8 или Z = −3,15) было бы редким событием, и исследователь может отклонить объяснение B. Если бы объяснение B было истинным, такое значение выборочной статистики было бы чрезвычайно маловероятным. Создается впечатление, что наша выборка, состоящая из 127 алкоголиков, проходящих лечение, извлечена из генеральной совокупности [illegible] чительно отличающейся от общества в целом по количеству прогулов.

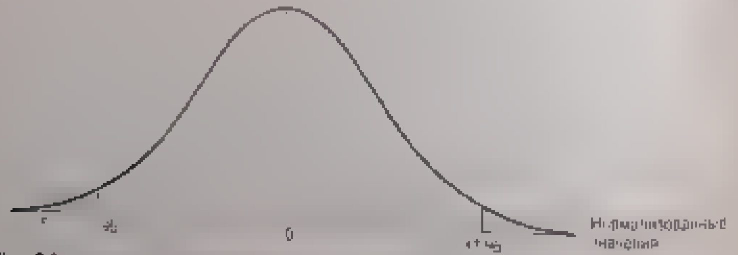


Рис. 8.3. [illegible]

[illegible] ...в редких случаях выборка может не являться [illegible] совокупности, из которой она была извлечена. [illegible] правильных ответов в резуль[illegible] принятия решений очень высока, но пока мы [illegible] генеральной совокупностью, этот про[illegible] относительно объяснены. В решении [illegible]

[illegible]

8.3. Пятишаговый алгоритм проверки статистических гипотез

В приведенном выше обсуждении нам удалось избежать всех формальных элементов и понятий, используемых при проверке гипотез. В этом разделе даны их правильные имена и представлен **пятишаговый алгоритм** процесса проверки гипотез.

Шаг 1. Формулировка допущений,

Шаг 2. Формулирование нулевой гипотезы,

Шаг 3. Подбор распределения выборочных статистик и задание критической области

Шаг 4. Вычисление статистики критерия,

Шаг 5. Принятие решения.

Рассмотрим каждый шаг отдельно, используя в качестве примера задачу из раздела 8.2.

Шаг 1. Формулировка допущений

[illegible] ...выборочном среднем мы должны... [illegible] должны предположить, что распределение выборочных средних [illegible] мальному закону. Мы будем называть все эти предположения [illegible] схема формирования выборки [illegible] ления "статистической моделью". Вы должны быть уверены [illegible] предположений.

Обычно мы будем указывать эти предположения в сокращенной [illegible] мер:

Модель: Выборка получена [illegible]
[illegible]
[illegible]

Шаг 2. Формулирование нулевой гипотезы (H_0)

[illegible] гипотеза — это формальное [illegible] [illegible] Такая форма нулевой гипо[illegible]

[illegible]

где μ означает среднее значение для совокупности алкоголиков, прошедших лечение. При проверке любой гипотезы нулевая гипотеза является центральным элементом, так как весь процесс нацелен на то, чтобы отклонить либо принять H_0.

Обычно исследователь считает, что наблюдаемое различие является значимым, поэтому он склонен к тому, чтобы отклонить нулевую гипотезу. На данном этапе [illegible] алгоритма мнение исследователя выражается в виде **альтернативной гипотезы** (H_1), которая представляет собой выражение, прямо противоположное нулевой гипотезе. Таким образом, цель исследователя при проверке гипотез часто заключается в том, чтобы собрать доказательства, которые говорили бы в пользу альтернативной гипотезы и на основании которых можно было бы отклонить нуле[illegible]

[illegible]

[illegible], в нашем [illegible] примера, имеющей среднее значение, не равн[illegible] ному значению.

$$[illegible]$$

где $\neq$ означает "не равно".

Это выражение, [illegible] [illegible] гласит, что выборка не [illegible] из генеральной [illegible] [illegible] 7,2, т.е. алкоголики, прошедшие курс лечения, [illegible]

[illegible]

[illegible] [illegible] о том, что по мнению исследо[illegible]

Шаг 3. Подбор распределения выборочных статистик и задание критической области

[illegible] [illegible] статистик является вероятностной ме[illegible] измерить любое [illegible]

статистики. Предположив, что нулевая гипотеза верна (и [illegible] путем такого [illegible] положения), мы можем вычислить значения для среднего и стандартного отклонения распределения выборочных статистик и, таким образом, измерить вероятность появления любого определенного значения выборочной статистики. Существует несколько различных распределений выборочных средних. Мы можем измерить [illegible] под кривой распределения, используя стандартную нормальную кривую (см. приложение А).

Критическая область состоит из площадей под кривой выборочного распределения, которые включают маловероятные значения выборочных статистик. [illegible] как проводить проверку гипотез, мы должны определить, что подразумевается под словом *маловероятные*. То есть, мы должны заранее указать те значения выборочных статистик, которые не приведут нас к отклонению H_0. Это решение [illegible] дает критическую область или **область отклонения**. Здесь использовано слово *область*, потому что мы описываем площадь под кривой распределения выборочных статистик, которая содержит маловероятные значения выборочных [illegible] [illegible] [illegible] примере эта площадь соответствует стандартизованному [illegible] значению, равному ±1,96, носящему название **критическое Z-значение**, которое графически изображено на рис. 8.2. Затемненная область называется критической областью. Любое значение выборочной статистики, для которого эквивалентное стандартизованное значение попадает в эту область (т.е. менее −1,96 или более +1,96), заставило бы нас отклонить нулевую гипотезу.

Условно размер критической области указывается как альфа (α) [illegible] всей площади, попадающей в критическую область. В примере, приведенном [illegible] значение альфа было равно 0,05. К другим распространенным значениям альфа относятся 0, [illegible] 0, 0,0[illegible] и [illegible]

Ниже в сокращенной форме указаны все решения, сделанные на этом шаге. Критическая область определяется в виде Z-значений, [illegible] [illegible]

[illegible]

Критическое значение Z (критическое) [illegible] ±1,96.

В качестве практического задания, касающегося [illegible] [illegible], см. задачу 8.1а.

Шаг 4. Вычисление статистики критерия

[illegible] оценить вероятность появления любого определенного [illegible] выбороч[illegible]

[illegible]

[illegible] экспериментальное Z-значение [illegible]

экспериментальное Z-значение равно [illegible] 3,1[illegible]. *[illegible] [illegible] касающегося вычисления экспериментальных Z-значений* [illegible]

Шаг 5. Принятие решения

[illegible] в процессе проверки статистической гипотезы является срав[illegible] критической области [illegible]

$$Z\text{ (критическое)} = \pm 1{,}96,$$

$$Z\text{ (экспериментальное)} = -3{,}15$$

и мы видели, что экспериментальное стандартизованное значение попало в критическую область (см. рис. 8.3). Мы приняли решение отклонить нулевую гипотезу и сделали вывод о том, что маловероятно, чтобы наблюдавшиеся нами различие в количестве прогулов было следствием простой случайности. С точки зрения количества прогулов, алкоголики, прошедшие курс лечения, отличаются от общества в целом.

Этот пятишаговый алгоритм будет служить в качестве основы для принятия решений в главах, посвященных проверке гипотез. Точный характер и методы формирования наших решений будут различными в различных ситуациях. Однако знание пятишагового алгоритма поможет вам в овладении этим материалом, так как он дает общую систему отсчета для проверки значимости.

8.4 Односторонний и двусторонний статистические критерии

Пятишаговый алгоритм проверки гипотез [illegible] дователи здесь весьма мало простор[illegible]

[illegible]

[illegible] гипотезе (H), которая [illegible] утверждает то, во что верит исследо[illegible] желает принять альтернативную гипо[illegible]

[illegible] в зависимости от связи между [illegible]

[illegible] на прогулы этой [illegible] форме, что "для всех алкоголиков [illegible] же уровень прогулов (7,2 дня), [illegible] Исследователь может считать [illegible]

[illegible] совокупности ниже, чем значение, указанное в нулевой гипотезе), либо *более высокий* уровень прогулов (среднее значение [illegible] указанное в нулевой гипотезе), или же он может быть не уверен в направлении различий.

Если исследователь не уверен в направлении, гипотеза исследования будет гласить лишь то, что среднее значение генеральной совокупности "не равно" значению, указанному в нулевой гипотезе. Альтернативная гипотеза, указанная в разделе 8.3 ($\mu \neq 7{,}2$), была такого вида. Для подтверждения такой альтерна[illegible] двусторонний статистический критерий [illegible] дователя интересовать как возможность того, что истинное значение генеральной совокупности выше значения, указанного в нулевой гипотезе, так и возможность того, что истинное значение генеральной совокупности ниже значения, указанного в нулевой гипотезе.

В других ситуациях исследователя могут интересовать различия в определенном направлении. Если направление различий можно предсказать, либо если исследователя интересуют только различия в одном направлении, можно использовать односторонний критерий. Односторонняя проверка может принимать одну из двух форм в зависимости от ожиданий исследователя относительно направления различий. Если исследователь считает, что истинное значение генеральной совокупности выше, чем значение, указанное в нулевой гипотезе, это мнение будет отражено в альтернативной гипотезе. Если бы в нашем примере мы предположили, что алкоголикам, прошедшим курс лечения, присущ более высокий уровень прогулов, чем остальному обществу (т.е. что для них среднее количество прогулов выше, чем 7,2), наша альтернативная гипотеза имела бы следующий вид.

[illegible]

где > означает "больше чем".

Если бы мы предположили, что алкоголикам, прошедшим курс лечения, [illegible] следующий вид.

H [illegible]

где < означает "меньше чем".

Часто односторонние критерии применяются [illegible] программ, разработанных для решения задач [illegible] пример, если программа лечения алкоголиков [illegible] ками, ее можно считать неудачной, как минимум [illegible] можно оценить программу, предназначенную для [illegible] статистические оценки должны использовать [illegible] ют о снижении безработицы. Если безработи[illegible] увеличивается, программа является [illegible] ресны для исследователей.

[illegible] …ритма выбор между односторонним и двусторон- [illegible]

правый хвост распределения [illegible] …что мы относим всю критическую [illegible] Если мы считаем, что значение параметра генеральной совокупности больше, чем значение, указанное в нулевой гипотезе (если H включает символ >), мы относим всю критическую область к правому хвосту распределения. Если мы считаем, что значение параметра генеральной совокупности меньше, чем значение, указанное в нулевой гипотезе (если H включает символ <), мы относим всю критическую область к левому хвосту распределения.

Например, в двусторонней проверке с альфа, равной 0,05, критическая область начинается с критического стандартизованного значения Z (критическое) = ±1,96. В случае одностороннего критерия при том же значении альфа критическое стандартизованное значение равно +1,65, если задан правый хвост распределения, и −1,65, если задан левый хвост распределения. Сравнение размещения критической области графически показано на рис. 8.4, а критические стандартизованные значения для наиболее распространенных значений альфа приведены в таблице 8.1 как для односторонних, так и для двусторонних критериев.

Таблица 8.1. Нахождение критических стандартизованных значений для одно- и двусторонних [illegible] по одной выборке

Альфа	Двустороннее значение	Одностороннее значение: Правый хвост	Одностороннее значение: Левый хвост
[illegible]	[illegible]	[illegible]	[illegible]

[illegible] …к среднему значению распределения [illegible] …С точки зрения статистики, односторонние критерии [illegible] …ресован лишь в различиях по [illegible] …студенты-социологи [illegible] …остальные студенты. Тест [illegible] При-

по "шкале утонченности" был проведен среди всех студентов и среди случайной выборки, состоящей из 100 студентов-социологов, и были получены следующие результаты:

Все студенты [illegible]

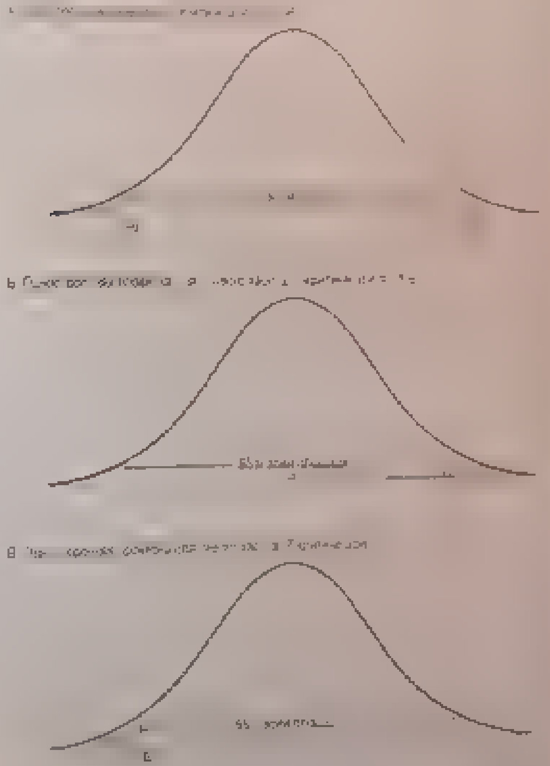


Рис. 8.4. [illegible] сравнение односторонних и [illegible] критериев [illegible] альфа [illegible] 0,05

[illegible] для того чтобы проверить гипотезу H_0, которая утверждает об отсутствии различий между [illegible] …ми студентами.

Шаг 1. Принятие предположений

Так как для обобщений выборочных [illegible] …ем, нам необходимо принять предположения [illegible] …яет выполнять измерения, которые будут [illegible]

[illegible] теорема, а значит мы можем [illegible] является нормальным

[illegible]

Шаг 2. Формулировка нулевой гипотезы (H_0)

[illegible] различия между студентами-социологами и все[illegible] время [illegible] поэтому оправдано приме[illegible]. Это можно выразить в следующем виде

$$H_0: \mu = 7{,}3$$

$$H_1: \mu > 7{,}3$$

Шаг 3. Подбор распределения выборочных статистик и задание критической области

[illegible] распределения выборочных [illegible] нормальным распределением [illegible] альфа [illegible] критическая область [illegible] выборка [illegible] в правом хвосте [illegible]

Распределение выборочных статистик = Z распределение,

α = 0,05 (односторонний критерий),

Z (критическое) = +1,65

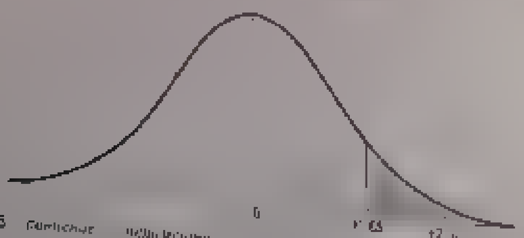


Рис. 8.5. [illegible] α = 0,05 [illegible]

Шаг 4. Вычисление статистики критерия

Z (экспериментальное) = [illegible]

Z (экспериментальное) = [illegible]

Z (экспериментальное) = +2,57

Шаг 5. Принятие решения

Сравнение экспериментального и критического стандартизованных значений

Z (критическое) = +1,65

Z (экспериментальное) = +2,57

Мы видим, что статистика критерия попадает в критическую область. Это значение графически представлено на рис. 8.5. Мы отклоним нулевую гипотезу, потому что если бы H_0 была истинной, такое различие было бы очень маловероятным. С точки зрения [illegible] имеет место значительное различие между студентами-социологами и студентами в целом. То, что нулевая гипотеза была отклонена, означает, что [illegible] *В качестве практического задания [illegible] см. задачи 8.2, 8.3, 8.6, 8.8 и 8.17*

8.5. Выбор значения альфа

Помимо того, что исследователь должен выбрать между [illegible] нужно выбрать значение альфа. Мы [illegible] значение альфа играет ключевую роль в проверке статистических гипотез. Когда мы [illegible] значения выборочной статистики меньше, [illegible] нулевую гипотезу, как [illegible] попадает в критическую область [illegible] на шаге 5.

Как принять рациональное решение относительно значения альфа? Помните, что [illegible] при проверке гипотез [illegible] гипотезы, или с отклонением нулевой гипотезы, [illegible]

[illegible] **ится ошибкой первого рода или альфа ошибкой.** Чтобы свести к м[illegible] [illegible] эту ошибку, следует использовать очень малые значения альфа [illegible]

[illegible] при обработке данных значения из этой области принудит нас отклонить нулевую гипотезу. Остальная область состоит из всех значений выборочных статистик, являющихся "нередкими". Чем меньше значение альфа, тем меньше критическая область и тем больше расстояние между средним значением распределения выборочных статистик и граничными точками критической области. В целях демонстрации сравните следующие значения альфа и критические стандартизованные значения для двух[illegible]

Если альфа равна	Двусторонняя критическая область проверки будет начинаться с критического [illegible]
0,10	±1,65
0,05	±1,96
0,01	±2,58
0,001	±3,29

По мере уменьшения значения альфа критическая область становится меньше и все больше удаляется от среднего значения распределения выборочных статистик, которому подчиняется выборка. Чем меньше значение альфа, тем сложнее отклонить нулевую гипотезу, и тем меньше вероятность ошибки первого рода [illegible]

[illegible] **[illegible]вание ошибки второго рода или** [illegible], в действительности, является [illegible]

[illegible] **принятием решений и** [illegible]дены для решения, которые могут быть [illegible] мы либо отклоняем нулевую гипотезу, [illegible] 8.2 [illegible]

[illegible] ("ОК"), а две из которых указывают на то, что была допущена ошибка.

Таблица 8.2. Принятие решения и нулевая гипотеза

H_0 на самом деле	Решение: Отклонить	Решение: Принять
[illegible]	[illegible]	[illegible]
Ложна	ОК	Ошибка второго рода (бета)

Два возможных значения приводят к отклонению нулевой гипотезы, которая и в самом деле является ложной, и к принятию нулевой гипотезы, которая в действительности является истинной. Цель любого научного исследования заключается в том, чтобы подтвердить истинные утверждения и отклонить ложные утверждения. Оставшиеся две комбинации представляют собой ошибки, то есть ситуации, которых мы, естественно, стремимся избежать. Если мы отклоним нулевую гипотезу, которая в действительности истинна, мы признаем ложным истинное утверждение. Подобным образом, если мы не отклоним нулевую гипотезу, которая в действительности является ложной, мы признаем истинным ложное утверждение. Очевидно, мы бы предпочли всегда оказываться в ячейках таблицы 8.2 с подписями "ОК" — всегда отклонять ложные и принимать истинные утверждения. Однако помните, что проверка гипотез всегда содержит элемент риска и что невозможно одновременно свести к минимуму вероятность ошибки и первого, и второго рода.

В конечном итоге это означает, что мы должны подбирать такое значение альфа, которое было бы балансом между двумя ошибками. Более высокие значения альфа сведут к минимуму вероятность ошибки второго рода (признание ложного утверждения истинным), и более низкие значения альфа сведут к минимуму вероятность ошибки первого рода (признание истинного утверждения ложным). В социологических исследованиях обычно нужно свести к минимуму ошибку первого рода, поэтому применяются более низкие значения альфа (0,05, 0,01, 0,01 и ниже). В частности, значение 0,05 считается общепринятым показателем значимого результата [illegible] которым [illegible] (например, 0,04, 0,027, 0,083). Исследователь несет ответственность за [illegible] значения альфа, которое выглядит наиболее разумным с точки зрения исследовательского проекта.

8.6. *t*-распределение Стьюдента

К этому моменту мы рассмотрели ли[illegible] мы сфокусировались на ситуациях, [illegible] [illegible] значением стандартного [illegible] в большинстве исследований значение [illegible] но для того, чтобы вычислить стандартн[illegible] образовать наше значение выборочн[illegible] и обозначить Z (экспериментальное) [illegible]

тик (шаг 4). Каким образом можно получить оправданное значение стандартного [illegible]

[illegible]

[illegible] дает адекватную оценку σ. Таким образом, для выборок большого размера на шаге 4 мы просто подставим в формуле вычисления Z (экспериментальное) s вместо σ и продолжим использовать стандартную нормальную кривую [illegible] распределения выборочных статистик.

[illegible] нахождения площади под кривой распределения выборочных статистик и [illegible] ления критической области следует использовать другое распределение, называемое *t*-распределением Стьюдента. Форма *t*-распределения изменяется в зависимости от размера выборки. Сравнение формы *t*-распределения и *Z*-распределения представлено на рис. 8.6. При маленьких размерах выборки кривая *t*-распределения является намного более пологой, чем кривая *Z*-распределения, но при увеличении размера выборки *t*-распределение все больше и больше начинает напоминать *Z* [illegible] идентичны. По мере того, как растет *N*, выборочное стандартное отклонение (*s*) ста[illegible] *t*-распределение [illegible]

[illegible] столбец "степени свободы". Как [illegible]

[illegible]

Третье отличие заключается в том, что элементы [illegible] значение *t* (критическое) [illegible]

[illegible]

[illegible] вы найдете значение 2,045. Таким образом, критическая область для этой проверки начинается в точке $t = \pm 2{,}045$.

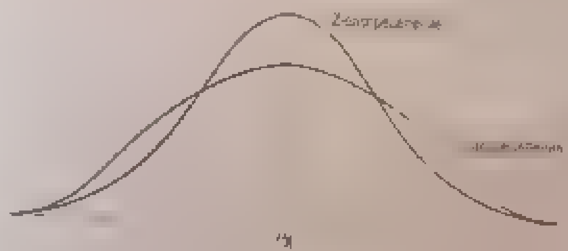


Рис. 8.6. [illegible] и *Z*-распределения

Давайте рассмотрим некоторые дополнительные свойства *t*-распределения. Во-первых, заметьте, что значение *t* (критическое), найденное нами выше, больше чем соответствующее ему значение *Z* (критическое), которое для двустороннего критерия при значении альфа, равном 0,05, равно ±1,96. Это связано с тем фактом, что *t*-распределение является более пологим, чем *Z*-распределение (см. рис. 8.6). При использовании *t*-распределения начальные точки критических областей будут находиться дальше от среднего значения выборочного распределения, и поэтому нулевую гипотезу будет сложнее отклонить. Более того, чем меньше размер выборки (чем меньше степеней свободы), тем более высоким будет значение *t* (экспериментальное), необходимое для отклонения H_0.

Во-вторых, просмотрим столбец для альфа при двустороннем критерии, равном 0,05. Обратите внимание на то, что при одной степени свободы *t* (критическое) равно ±12,706, и что значение *t* (критическое) сжимается по мере того, как растет число степеней свободы. При числе степеней свободы, превышающем 120, значение *t* (критическое) равно соответствующему значению *Z* (критическое), т.е. ±1,96. По мере роста размера выборки *t*-распределение становится все более похожим на *Z*-распределение [illegible] не станут идентичными.[4]

Чтобы более подробно продемонстрировать *t*-распределение, [illegible] обратите внимание на то, что с точки зрения [illegible] данные с использованием *t*-значений, возникают [illegible]

[illegible]

[illegible] выборочных статистик следует использовать [illegible] число степеней свободы следует вычислить перед определением [illegible] области, обозначенной через t (критическое). На шаге [illegible] статистики критерия, t (экспериментальное), использу[illegible] формула. В отличие от формулы для вычисления Z (эксперименталь[illegible]) [illegible] σ, и $N-1$ заменяет N.

[illegible]

Формула 8.2 [illegible] $= \dfrac{\bar{X}-\mu}{s/\sqrt{N-1}}$

Исследователь хочет узнать, отличаются ли студенты, живущие в пригородах, от [illegible] выборку, состоящую из 30 студентов, живущих в пригородах, и узнал от секретаря, что средний балл для всех студентов равен 2,50 ($\mu = 2,50$), но стандартное отклонение генеральной совокупности (σ) неизвестно. Ниже приведены данные выборки. Является ли выборка частью генеральной совокупности, имеющей сред[illegible]?

Все студенты	Студенты, живущие в пригородах
$\mu = 2{,}50$ (= μ_X)	[illegible] 2,78
$\sigma = ?$	$s =$ [illegible]
	$N =$ [illegible]

Шаг 1. Выдвижение предположений

Модель. Формирование случайной выборки
Интервальная/относительная шкала измерений
Нормальное распределение выборочных статистик

Шаг 2. Формулировка нулевой гипотезы (H_0)

H_0: $\mu = 2{,}50$
[illegible]

[illegible]

Шаг 3. Подбор распределения выборочных статистик и задание критической области

[illegible]

Рис. 8.7
Распределение [illegible] статистик [illegible] сравнение [illegible] критерия [illegible]

Шаг 4. Вычисление статистики критерия

$$t\text{ (экспериментальное)} = \frac{\bar{X}-\mu}{s/\sqrt{N-1}}$$

t (экспериментальное) = [illegible]

[illegible]

Шаг 5. Принятие решения

Статистика критерия не попадает в критическую область. Поэтому исследователь не может отклонить H_0. Различие между средним значением выборки (2,78) и средним значением генеральной совокупности (2,50) не превышает по величине различий, которые могли бы возникнуть вследствие одной лишь случайности. Статистика критерия и критическая область представлены на рис. 8.7.

Подводя итоги, скажем, что при проверке средних значений, полученных для одной выборки, мы должны сделать выбор теоретического распределения, которое мы будем использовать для установления критической области. Выбор достаточно прост. Если известно стандартное отклонение (σ) или выборка имеет большой раз[illegible]. Если значение σ неизвестно и выборка имеет небольшой размер, следует использовать t-распределение (приведен[illegible]) *[illegible] задания, касающиеся использования t-распределения для провер[illegible]*

Практический пример 8.1

[illegible]

[illegible] мере от среднего срока лишения свободы [illegible] по стране? Ниже представ[illegible]

[illegible]

$s = 3$

$N = 152$

[illegible]

[illegible] гипотеза утверждает, что выборка извлечена из генеральной совокупно[illegible]

[illegible] критерия, Z (экспериментальное) будет равна

$$Z\text{ (экспериментальное)} = \frac{\bar{X} - \mu}{s/\sqrt{N-1}}$$

$$Z\text{ (экспериментальное)} = \frac{[illegible]}{3{,}7/\sqrt{152-1}}$$

$$Z\text{ (экспериментальное)} = \frac{[illegible]}{3{,}7/\sqrt{151}}$$

$$Z\text{ (экспериментальное)} = \frac{-1{,}40}{0{,}30}$$

$$Z\text{ (экспериментальное)} = -4{,}67$$

Если альфа имеет значение 0,05 [illegible] Z (критическое) $= \pm 1{,}96$. Наше [illegible] гипотезу. Различие между сро[illegible] целом, является статистически значимым. Раз[illegible]

8.7 Проверка гипотез для пропорции одной выборки (большие выборки)

[illegible]

качестве статистики критерия выборочной пропорции P_s [illegible] выборки. Как мы увидим дальше, общие процедуры проверки пропорций по одной выборке идентичны используемым при проверке средних значений. Главным вопросом все еще является вопрос "Имеет ли генеральная совокупность, из которой была извлечена выборка, данную характеристику?" Мы все еще выполняем проверку, основываясь на предположении, что нулевая гипотеза истинна, и мы все еще оцениваем вероятность возникновения случайного значения выборки относи[illegible] предположить, что H_0 истинна), тогда мы отклоняем H_0.

После этого [illegible] тезы используются другие символы, несмотря на то, что она все еще представляет собой утверждение об "отсутствии различий".

На шаге 3 мы для поиска значений площади под кривой распределения выборочных статистик и определения местонахождения критической области будем использовать только стандартную нормальную кривую (Z-распределение). Это приемлемо в тех случаях, когда выборка имеет большой размер. В этой книге мы не будем рассматривать проверку гипотез для пропорций выборок небольшого размера.

Практический пример 8.2

В случайной выборке, составленной из жителей наиболее обеспеченного района [illegible] во время последних президентских выборов голосовали [illegible] за республиканцев. Можно ли сделать вывод, что среди жителей наиболее обеспеченного района значительно больше людей голосовали за республиканцев? Ниже приведена информация, необходимая для проверки нулевой гипотезы, выраженная в виде долей.

Наиболее обеспеченный район [illegible]

$P_s =$ [illegible] $P_u =$ [illegible]

$N = 103$

Наша нулевая гипотеза имеет следующий вид.

$H_0: P_u =$ [illegible]

Другими словами, нулевая [illegible] генеральной совокупности, 66% [illegible]

Тестовый статистический пок[illegible] Z [illegible]

$$Z\text{ (экспериментальное)} = \frac{P_s - P_u}{\sqrt{P_u(1-P_u)/N}}$$

$$Z\text{ (экспериментальное)} = \frac{[illegible]\ 0{,}66}{\sqrt{(0{,}66)(1-0{,}66)/103}},$$

$$Z\text{ (экспериментальное)} = \frac{0{,}[illegible]00}{0{,}047}$$

$$Z\text{ (экспериментальное)} = 2{,}[illegible]3$$

Так как нас интересует вопрос, действительно ли "среди жителей наиболее обеспеченного района значительно больше людей голосовали за республиканцев", требуется односторонняя проверка гипотез. При значении альфа, равном 0,05, наша критическая область начнется с Z (критическое) = +1,65. Критическая область располагается в правом хвосте, так как мы предсказываем, что выборка извлечена из генеральной совокупности, в которой за республиканцев проголосовало больше людей, чем в целом по городу.
Так как наблюдаемое нами стандартизованное значение равно 2,13, следует отклонить нулевую гипотезу. Различие между жителями богатого района и жителями города в целом является статистически значимым и имеет предсказанное направление. Среди жителей богатого района значительно больше людей голосовали за республиканцев.

На шаге 4, где происходит вычисление статистики критерия, формула имеет тот [illegible] есть статистика критерия, Z (экспериментальное), равно разности меж- [illegible]

Формула 8.3 $Z\text{ (экспериментальное)} = \frac{P_s - P_u}{\sqrt{P_u(1-P_u)/N}}$

Шаг 5, принятие решения, остается таким же, как и прежде. Если статистика критерия, Z (экспериментальное), попадает в критическую область, обозначен[illegible] через Z (критическое), следует отклонить H [illegible]

[illegible]

Шаг 1. Выдвижение предположений

Модель. Формирование случайной выборки
Номинальная шкала
Нормальное распределение выборочных статистик

Шаг 2. Формулирование нулевой гипотезы (H_0)

Как говорилось выше, вопрос исследования касается лишь существования различий между выборочной пропорцией и пропорцией в генеральной совокупности. Так как не предсказывалось направление различий, следует использовать двусторонний критерий.

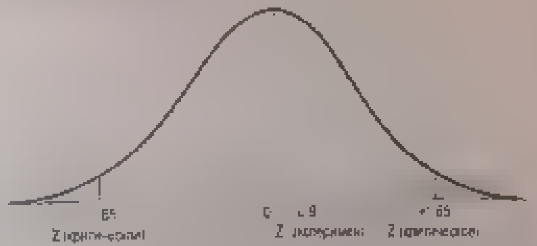
$$H_0: P_u = 0{,}39$$

$$(H_1: P_u \neq 0{,}39)$$

Шаг 3. Подбор распределения выборочных статистик и задание критической области

Распределение выборочных статистик = Z-распределение,
α = 0,10 (двусторонний критерий),
Z (критическое) = ±1,65

Рис. 8.8.
Распределение, представляющее сравнение экспериментального и критического Z-значений (α = 0,10, двусторонний критерий)



Шаг 4. Вычисление статистики критерия

$$Z\text{ (экспериментальное)} = \frac{P_s - P_u}{\sqrt{P_u(1-P_u)/N}},$$

$$Z\text{ (экспериментальное)} = \frac{0{,}43-0{,}39}{\sqrt{(0{,}39)(0{,}61)/122}};$$

$$Z\text{ (экспериментальное)} = +0{,}9[illegible]$$

Шаг 5. Принятие решения

Статистика критерия, Z (экспериментальное), [illegible] Поэтому мы не можем отклонить H_0 [illegible] торых главой семьи является жена, не существует [illegible]

[illegible] между значениями [illegible] и Z (экспериментальное) представлены на [illegible] критическая область [illegible]

Заключение

1 В этой главе были представлены все основные понятия и методы проверки гипотез. [illegible] вида об "отсутствии различий" между средними значениями выборки и пропорциями. В обоих случаях главным вопросом является вопрос "Имеет ли генеральная совокупность, из которой была извлечена выборка, данную характеристику?"

2 Все проверки статистических гипотез включают в себя вычисление вероятности наблюдаемого значения выборочной статистики, при условии, что нулевая гипотеза истинна. Если значениям выборочной статистики соответствуют низкие значения вероятности, мы отклоняем нулевую гипотезу. Обычно мы будем стремиться к тому, чтобы отклонить нулевую гипотезу и таким образом принять альтернативную.

3 Пятишаговый алгоритм будет нашей общей схемой принятия решений во всех главах, посвященных проверке гипотез. [illegible] (2) формулирова[illegible]

[illegible]

4 [illegible] требуется [illegible] предсказать направление изменения [illegible] При проверке гипотез существует [illegible] ошибка) связана с принятием [illegible] является ложной. Вероятности [illegible] При выборе значения альфа мы [illegible] этих двух видов.

5 [illegible] средних, если стандартное отклонение генеральной [illegible] и размер выборки мал [illegible] применять t-распределение.

6 [illegible] применением пятишагового [illegible] главные различия [illegible] измерения [illegible] статистики [illegible]

7 Если вам все еще [illegible] пусть это вас не тревожит [illegible] достаточно сложного материала [illegible] проверки статистических гипотез. Кроме того [illegible] после того, как [illegible] нулевая гипотеза [illegible] с помощью распределения [illegible] деление выборочных статистик).

Сводка формул

Выборочные средние одной выборки, большие выборки

8.1
$$Z \text{ (экспериментальное)} = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

Выборочные средние одной выборки при малых выборках и неизвестном стандартном отклонении генеральной совокупности

8.2
$$t \text{ (экспериментальное)} = \frac{\bar{X} - \mu}{s/\sqrt{N-1}}$$

Выборочные пропорции одной выборки, большие выборки

8.3
$$Z \text{ (экспериментальное)} = \frac{P_s - P_u}{\sqrt{P_u(1-P_u)/N}}$$

Словарь

Альтернативная гипотеза (H_1). Утверждение, противоположное нулевой гипотезе. В контексте проверки значимости по одной выборке эта гипотеза предполагает, что генеральная совокупность, из которой была извлечена выборка, не имеет определенной характеристики или значения.

Двусторонний критерий. Критерий для проверки статистических гипотез, применяемый в случаях, когда (1) нельзя предсказать направление различий, (2) исследователя интересуют значения, которые находятся в обоих хвостах распределения выборочных статистик.

Критическая область (область отклонения нулевой гипотезы). Область под кривой выборочного распределения, которая перед проведением проверки определяется, как содержащая такие значения выборочных статистик, [illegible] которых при условии истинности нулевой гипотезы маловероятно.

Нулевая гипотеза (H_0). Утверждение об "отсутствии различий". В контексте проверки значимости в случае одной выборки эта гипотеза [illegible] генеральная совокупность, из которой была [illegible] ленную характеристику или заданное значение.

Односторонний критерий. Критерий для проверки статистических гипотез, применяемый в случаях, когда (1) можно предсказать [illegible]

исследователя интересуют только те значения, которые находятся в определ[illegible] [illegible] части распределения выборочных статистик.

[illegible] Вероятность того, что не будет отклонена ну[illegible]

Ошибка второго рода (бета-ошибка) [illegible]

Ошибка первого рода (альфа-ошибка) [illegible]

Проверка значимости [illegible]

Проверка статистических гипотез [illegible]

[illegible] [illegible]

Статистика критерия [illegible] Z [illegible]

Уровень альфа (α) [illegible]

Критическ[illegible] [illegible]

t-распределение Стьюдента [illegible]

[illegible]

Z (критическое) [illegible]

Z [illegible]

Мультимедийные ресурсы

[illegible] Center. Virtual Society (Виртуальное сообщество [illegible]

[illegible]

- Обзор основ математики
- Обучающие карточки
- Дополнительные задачи с ответами
- Обзор статистики
- [illegible]
- Таблица случайных чисел
- Примеры и упражнения с применением пакетов статистической обработки данных MicroCase и SPSS:
- Функциональные диаграммы "Find the test"
- Проверка гипотез для переменных, измеряемых по порядковой шкале

Задачи

8.1. а. Для каждой ситуации найдите Z (критическое)

Альфа	Вид	Z (критическое)
[illegible]	[illegible]	
0[illegible]	[illegible]	
[illegible]	[illegible]	
	[illegible]	
[illegible]	[illegible]	

б. Для каждой ситуации найдите t (критическое)

Альфа	Вид	N	t (критическое)
[illegible]	[illegible]	3	
[illegible]	[illegible]	24	
[illegible]	[illegible]		
[illegible]	[illegible]	[illegible]	
[illegible]	[illegible]	[illegible]	

в. Вычислите подходящую статистику критерия (Z [illegible]

[illegible]

1. $\mu = 2{,}40$ [illegible]
2. [illegible]
3. [illegible] N = [illegible]
4. P_u = [illegible]
5. P_u = [illegible]

8.2. SOC

а. [illegible] вечеринок в месяц со стандартным отклонением 0,3. [illegible] вывод, что студенты-социологи значительно отличают[illegible] в целом? (*[illegible] формулировка задачи подразумевает применение двустороннего критерия. Это означает, что альтернативная гипотеза* [illegible] *как H_1: $\mu \neq 3,7$, и что критическая* [illegible] *Значения Z (критическое) можно найти в таблице 8.1* [illegible]

б. Если [illegible] вопрос [illegible] значительно большее количество вечеринок?", что бы изменилось [illegible] проверке, проводимой в пункте 8.2а? *Подсказка: такая формулировка подразумевает применение двустороннего критерия для проверки значимости. Как изменится альтернативная гипотеза исследования? Чему будет равно значение Z (критическое), если применять значение альфа, использовавшееся в задаче 8.2а?*

8.3. SW

а. В национальных масштабах опыт работы социальных работников составляет в среднем 10,2 лет. В случайной выборке, состоящей из 203 социальных работников города Шинбоун, опыт работы составил в среднем всего 8,7 лет со стандартным отклонением 0,52. Можно ли сделать заключение, что социальные работники города Шинбоун значительно менее опытны? *Обратите внимание на формулировку альтернативной гипотезы. В ситуациях, подобных этой, может быть обоснованным применение одностороннего критерия для проверки значимости. Если вы выберете односторонний критерий, какой вид примет альтернативная гипотеза, и где начнется критическая область?*

б. В той же самой выборке социальных работников средняя годовая зарплата составляет $25782 со стандартным отклонением $622. Является ли эта цифра значительно более высокой, чем средняя зарплата социальных работников по стране, составляющая $24500? *Формулировка альтернативной гипотезы подразумевает применение одностороннего критерия для проверки значимости. Какой вид должна принять альтернативная гипотеза, и где начнется критическая область?*

8.4. SOC

Средний балл на вступительных экзаменах в колледж по стране (устный экзамен) [illegible] со стандартным отклонением 95. В случайной выборке, состоящей из 152 первокурсников, средний балл составил 502. Является ли статистически значимым это различие?

8.5. SOC

[illegible] Америки? [illegible]

8.6. SOC

Средняя зарплата в выборке, состоящей из 35 работников подразделения Overk [illegible] компании [illegible] составляет $24[illegible] [illegible] средняя зарплата всех работников [illegible] составляет $24[illegible] [illegible] $[illegible] Можно ли сказать [illegible] Overk [illegible] зарабатывают больше [illegible] критерий.

8.7. GER

а. [illegible] 6,2 [illegible] В результате [illegible] случайной выборки [illegible] что они смотрят телевизор в среднем 5,9 часов в день со стандартным отклонением 0,7. Является ли это различие статистически значимым?

б. На [illegible] той же самой выборки [illegible] в среднем [illegible] со стандартным отклонением 0,[illegible] [illegible] 1,7. Является ли это различие статистически значимым?

8.8. SOC

[illegible] несколько сотен "хронических [illegible]" [illegible] в программе альтернативного образования. [illegible] сформирована выборка из 35 участников для сравнения с другими учащимися системы.

а. Имеет ли программа положительный результат с точки зрения среднего балла?

Средний балл по системе	Средний балл по программе
$\mu = 2,47$	$\bar{X} = 2,55$
	$s = [illegible]$
	$N = 35$

б. С точки зрения числа прогулов (количества пропущенных дней в году), что можно сказать об успехе программы?

По системе	По программе
$\mu = 6,13$	$\bar{X} = 4,[illegible]$
	$s = [illegible]$
	$N = 35$

в. Была ли программа успешной с точки зрения оценок за стандартные тесты по математике и чтению?

Тест по математике: По системе	По программе	Тест по чтению: По системе	По программе
$\mu = 103$	$\bar{X} = 106$	$\mu = 110$	$\bar{X} = 113$
	$s = 2,0$		$s = 2,0$
	$N = 35$		$N = 35$

Подсказка: обратите внимание на формулировку гипотезы исследования. Обоснованно ли применение одностороннего критерия? [illegible]

[illegible] правыми не отличающися от учащихся по [illegible] [illegible] программы были ниже? [illegible]

8.9. SOC

[illegible]

8.10. [illegible]

В среднем по стране налог на собственность составил $ [illegible] на душу населения [illegible] тоже, этот показатель составил $98 со стандартным отклонением $5. Является [illegible]

8.11. [illegible]

Исследование показало, что каждый год 10% населения становится жертвами имущественных преступлений. В случайной выборке из 327 пожилых жителей (в возрасте 65 лет и более) уровень виктимизации составил 14%. Можно [illegible]

8.12. CJ

Случайная выборка из 113 осужденных насильников, отбывающих наказание [illegible] прошли программу, нацеленную на то, чтобы изменить их отношение к жен[illegible]

8.13. PS

[illegible] проведенном на территории штата, 55% жителей [illegible] предложения о выпуске лотерей. В случайной выборке [illegible] отклоняют это предложение. [illegible]

8.14. CJ

[illegible] собрал случайную выборку из всех [illegible] в которых было совершено [illegible]

уровнем раскрытия преступлений по штату? Напишите [illegible] предложений, объясняющих ваше решение.

8.15. SOC/SW

[illegible] файл с информацией о 317 семьях [illegible] случайную выборку из семей города, в которых на протяже[illegible] имеет место жестокое обращение с детьми. Ниже [illegible] характеристики выборки вместе с соот[illegible]

а. [illegible]

Город	Выборка
$P_u = 0{,}63$	$P_s = 0{,}69$

б. Размер семьи (пропорция семей, имеющих более четырех детей).

Город	Выборка
$P_u = 0{,}21$	$P_s = 0{,}26$

в. Профессия матери (пропорция матерей, работающих вне дома).

Город	Выборка
$P_u = 0{,}5$[illegible]	$P_s = 0$[illegible]

г. Отношения с родственниками (пропорция семей, имеющих контакт с родственниками по меньшей мере раз в неделю).

Город	Выборка
$P_u = 0{,}82$	$P_s = 0{,}43$

д. Образованность отца (средняя продолжительность формального образования[illegible]).

Город	Выборка
$\mu = 12{,}3$	$\bar{X} = 12{,}5$
	$s = 1{,}7$

е. Стабильность занятости отца (среднее количество лет, в течение которых отец работает на текущей работе).

Город	Выборка
[illegible]	[illegible]
	$s = 0{,}5$

8.16. SW

[illegible] рования программы снижен[illegible]

[illegible] состоящей из 323 подростков-юношей уровень безработицы [illegible] статистически значимым? [illegible] так, как бы вы объяснили их в агентстве по [illegible]

8.17. [illegible]

[illegible] ронний критерий? Почему? Что он может сказать в служебной записке союз[illegible]

8.18. Следующие вопросы касаются основных принципов и понятий статистических выводов. Порядок вопросов приблизительно соответствует пятишаговому [illegible]

а. [illegible]

в. Объясните роль распределения выборочных статистик в проверке гипотез.

г. Нулевая гипотеза представляет собой предположение относительно действительности, с помощью которого становится возможным проверять значи[illegible]

[illegible]

9 Проверка статистических гипотез II. Случай двух выборок

Цели изучения [illegible]

1. [illegible] двух [illegible]
2. Объяснить [illegible] с двумя выборками
3. Объяснить, что такое независимая случайная выборка
4. [illegible]
5. [illegible] тезы. Объяснить различия между статистической значимостью и важностью.

9.1 Введение

[illegible]

Главным вопросом, возникающим при [illegible] с двумя выборками, является вопрос "[illegible] таточно большим для того, чтобы мы [illegible] тью ошибки), что генеральные [illegible] различаются?" Таким образом [illegible] сительно поддержки контроля [illegible]

чин и женщин, мы можем утверждать, что различия между выборками не возникли вследствие простой случайности [illegible]

В этой главе мы рассмотрим проверку значимости различий между выборочными средними и между выборочными пропорциями. В обоих случаях мы будем [illegible] для организации принятия решений. Хотя общие алгоритмы процесса проверки гипотез здесь очень похожи на алгоритм, применяемый при наблюдениях с одной выборкой, имеют место некоторые важные различия, которые мы и рассмотрим в этой главе.

9.2. Проверка статистических гипотез относительно выборочных средних (большие выборки)

Одно из важнейших различий между случаем одной выборки и случаем с двух выборок возникает на шаге 1 пятишагового алгоритма. В случаях с одной выборкой мы предполагали, что выборка была получена на основе метода случайного отбора (следуя принципу МРВО). Как вы помните, это подразумевает, что у всех наблюдений в генеральной совокупности должны быть равные шансы на то, чтобы попасть в выборку. В ситуации с двумя выборками выборки должны быть сформированы не только используя метод случайного отбора, но еще и независимо друг от друга. Это требование удовлетворяется в случаях, когда выбор конкретного объекта для одной выборки не оказывает влияния на вероятность того, что некоторый другой объект будет включен в другую выборку. В нашем примере, рассматривающем различия между полами в вопросе о поддержке контроля за продажей оружия, это означает, что попадание в выборку конкретного мужчины не оказывает влияние на вероятность попадания в выборку любой конкретной женщины. Это новое [illegible] шаге 1 называется получение выборки методом независимого случайного отбора [illegible]

[illegible] вание получена используя метод случайного отбора [illegible] МРВ [illegible] разделить ее элементы на отдельные группы (например, на мужчин и на женщин) [illegible]

[illegible] принимает нулевая гипотеза. Нулевая гипотеза [illegible] ет собой утверждение об "отсутствии различий". Однако теперь [illegible] том, что генеральная совокупность, из которой была [illegible] одну характеристику, она утверждает, что между дву [illegible] заношением оружия"). Если [illegible] нулевую гипотезу, утвержда [illegible] совокупностями, можно откло [illegible] что между ними нет различий и

[illegible] возможные разности между [illegible] ля за продажей оружия среди мужчин и среди женщин. Если нулевая гипотеза истинна, и между мужчинами и женщинами нет различий во взглядах относительно контроля оружия, тогда разность между средними значениями генеральных совокупностей была бы равна нулю, среднее значение распределения выборочных статистик было бы равно нулю, и, в большинстве своем, различия между средними значениями выборок были бы малыми. Чем больше различия между средними значениями выборки, тем дальше значение выборочной статистики (разность между двумя средними значениями выборок) будет находиться от среднего значения из выборочных средних распределения (от нуля), и тем больше вероятность того, что разность будет отражать действительное различие между генеральными совокупностями, представленными выборками.

Чтобы проиллюстрировать процедуру проверки выборочных средних, предположим, что исследователь имеет доступ к выборке, репрезентативной на национальном уровне, и что респонденты выборки дали ответы на вопросы анкеты в соответствии со шкалой для измерения отношения к контролю за продажей оружия. Выборка разделена по полу, и статистика выборки вычисляется для мужчин и для женщин. Предполагая, что данные относятся к интервальной/относительной шкале, можно провести проверку значимости различий между средними значениями выборок.

Если выборка имеет большой размер (то есть, если суммарное количество наблюдений в двух выборках превышает 100), распределение разностей средних значений выборок будет нормальным и для установления критических областей можно использовать нормальную кривую (приложение А). Статистика критерия Z (эксперим[illegible] [illegible] стики (разность между средними значениями выборок, минус среднее значение [illegible] формула [illegible] мы имеем дело с распределением разностей между средними значениями выборок.

Формула 9.1 Z [illegible]

где: $(\bar{X}_1 - \bar{X}_2)$ — разность средних значений выборок,

$(\mu_1 - \mu_2)$ — разность средних значений генеральных совокупностей,

$\sigma_{\bar{X}-\bar{X}}$ — стандартное отклонение распределения значений разностей средних значений выборок.

[illegible] так как мы [illegible]

В [illegible] H_0: [illegible] $= \mu$ [illegible] нулевая гипотеза (формулируемая [illegible]) [illegible] основываются на предположении [illegible]. Если [illegible] [illegible]. В [illegible] формула [illegible] для вычисления статистики критерия, имеет следующий вид:

Формула 9.2 [illegible]

Для больших выборок стандартное отклонение распределения разностей выборочных показателей вычисляется следующим образом:

Формула 9.3 [illegible]

Так как лишь в редких случаях мы будем знать значения стандартных отклонений генеральных совокупностей (σ_1 и σ_2), чтобы их оценить, нам следует использовать стандартные отклонения выборок, в которые введены надлежащие поправки на смещения. В формуле 9.4 приведено уравнение, которое используется для оценки стандартного отклонения выборочного распределения в данной ситуации. Этот показатель называется обобщенной оценкой, так как он объединяет информацию из обеих выборок.

Формула 9.4 [illegible]

Ниже приведены значения выборочных статистик в пользу контроля за продажей оружия, и теперь мы можем выполнить проверку значимости.

Выборка 1	Выборка 2
[illegible]	[illegible]
$\bar{X}_1 = 6{,}2$	$\bar{X}_2 = 6{,}5$
[illegible]	[illegible]
[illegible]	[illegible]

[illegible] показателей выборок, что для мужчин среднее значение [illegible] контроля за ношением оружия ниже. Это означает, что [illegible] достаточно большим, чтобы подтвердить [illegible] не вследствие простой случайности, а скорее всего [illegible] различие между мужчинами и женщинами по этому [illegible]

Шаг 1. Принятие предположений

Обратите внимание на то, что хотя теперь мы предполагаем, что случайные выборки являются независимыми, остающиеся предположения [illegible] ми же, как и в случае с одной выборкой.

Модель: Независимые случайные выборки
Интервальная/относительная шкала
Нормальное распределение выборочных статистик

Шаг 2. Формулирование нулевой гипотезы (H_0)

Нулевая гипотеза утверждает, что *генеральные совокупности*, представленные выборками, не отличаются по этой переменной. Так как направление различий не указано, требуется применение двустороннего критерия, что отражено в альтернативной гипотезе.

$$H_0: [illegible]$$
$$H_1: [illegible]$$

Шаг 3. Подбор распределения выборочных статистик и установление критической области

При больших размерах выборок для нахождения площадей под кривой выборочного распределения выборочных статистик и установления критической области можно использовать Z-распределение. Альфа будет иметь значение 0,05.

Распределение выборочных статистик = Z-распределение,
альфа = 0,05, двусторонняя,
Z (критическое) = $\pm 1{,}96$.

Шаг 4. Вычисление статистики критерия

Так как стандартное отклонение генеральных совокупностей неизвестны, для оценки стандартного отклонения распределения выборочных статистик можно применить формулу 9.4. Затем это значение можно подставить в формулу 9.2, чтобы [illegible] Z [illegible]

$$\sigma_{\bar{X}-\bar{X}} = [illegible]$$

$$\sigma_{\bar{X}-\bar{X}} = [illegible]$$

$$\sigma_{\bar{X}-\bar{X}} = [illegible]$$

$$\sigma_{\bar{X}-\bar{X}} = [illegible]$$

$$\sigma_{\bar{X}-\bar{X}} = [illegible]$$

$$Z(\text{экспериментальное}) = \frac{[illegible]}{\sigma}$$

$$Z(\text{экспериментальное}) = \frac{[illegible]}{0{,}107}$$

$$Z(\text{экспериментальное}) = \frac{[illegible]}{0{,}107}$$

Z(экспериментальное) = [illegible]

Шаг 5. Принятие решения

Сравнение статистики критерия и критической области

Z(экспериментальное) = [illegible]

Z(критическое) = ±1,96

Мы видим, что наблюдаемое значение статистики критерия явно попадает в критическую область. Это указывает на то, что возникновение различий, равного -0,3 (6,2 − 6,5), между средними значениями выборок при условии истинности нулевой гипотезы является маловероятным. Нулевую гипотезу, утверждающую об отсутствии различий, можно отклонить, что подтверждает ту точку зрения, что между мужчинами и женщинами существуют различия во мнениях относительно поддержки контроля оружия. Вероятность того, что решение отклонить нулевую гипотезу является неверным, равна всего лишь 0,05 (значение альфа).

Обратите внимание на то, что значение Z(экспериментальное) является отрицательным, а это свидетельствует о том, что [illegible] за понением оружия значительно ниже [illegible] как перед статистикой критерия отражает наше произвольное решение отнести мужчин к выборке 1, а жен[illegible] [illegible]

9.3 Проверка гипотез относительно выборочных средних (небольшие выборки)

[illegible] и, таким образом, для выяв-

ления маловероятных значений выборочных статистик. Для того чтобы применить t-распределение для проверки двух средних значений выборок, нам нужно выполнить одно дополнительное вычисление и сделать дополнительное предположение. Следует вычислить число степеней свободы величины, требуемой для правильного использования таблицы t-значений (приложение Б). В случае с двумя выборками число степеней свободы равно $N_1 + N_2 - 2$.

Дополнительное предположение является более сложным вопросом. При небольших выборках, для того чтобы обосновать предположение о нормальности распределения выборочных статистик и для того чтобы найти объединенную оценку стандартного отклонения их распределения, мы должны предположить, что дисперсии генеральных совокупностей, представляющих интерес, равны ($\sigma_1^2 = \sigma_2^2$). Предположение о равенстве дисперсий генеральных совокупностей можно проверить с помощью метода статистических выводов, называемого дисперсионного анализа или ANOVA (см. главу 10). Однако пока мы просто предположим, что имеет место равенство дисперсий генеральных совокупностей без формальной проверки. Пока размеры выборок приблизительно одинаковы, это предположение можно считать обоснованным.

Чтобы проиллюстрировать эту процедуру, предположим, что исследователь [illegible] тает, что семьи, проживающие в центре города, значительно больше по размеру (количеству детей), чем семьи, живущие в пригородах. Были сформированы случайные выборки из жителей обоих районов, и вычислены статистические показатели выборок.

Выборка 1 Пригород	Выборка 2 Центр города
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

Данные выборок выявляют различия в предполагаемом направлении. Значимость наблюдаемого различия может быть проверена с помощью пятишагового алгоритма.

Шаг 1. Формулирование предположений

Размер выборки небольшой, и стандартное отклонение генеральной совокупности [illegible] имеет место равенство дисперсий генеральных совокупностей.

Модель: Независимые случайные выборки
[illegible]
Дисперсии генеральных [illegible]
Нормальное распределение выборочных статистик

Шаг 2. Формулирование нулевой гипотезы

Так как было предположено направление (семьи, проживающие [illegible], больше по размеру), будем использовать односторонний [illegible] [illegible]

(H_1: $\mu_1 < \mu_2$).

Шаг 3. Подбор распределения и выборочных статистик установление критической области

При малых размерах выборок для установления критической области можно использовать t-распределение. Альфа будет иметь значение 0,05, и будет использован односторонний критерий.

Распределение выборочных статистик = t-распределение.

Альфа = 0,05, односторонний

Число степеней свободы = $N_1 + N_2 - 2 = 42 + 37$ [illegible]

Обратите внимание на то, что критическая область отнесена в левый хвост выборочного распределения, в соответствии с направлением, определенным в H_1.

Шаг 4. Вычисление статистики критерия

При небольших размерах выборок для нахождения обобщенной оценки стандартного отклонения распределения выборочных статистик используется другая формула (формула 9.5). Это значение затем подставляется напрямую в знаменатель формулы для нахождения t (экспериментальное), представленной в виде формулы 9.6.

Формула 9.5

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}}\sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{(42)(0,63)^2 + (37)(0,95)^2}{42+37-2}}\sqrt{\frac{42+37}{(42)(37)}}$$

$$= \sqrt{\frac{50,06}{77}}\sqrt{\frac{79}{554}},$$

$$\sigma_{\bar{X}-\bar{X}} = (0,81)(0,23)$$

Формула 9.6

Шаг 5. Принятие решения

Сравним значения статистики критерия и значения границы критической области:

t (экспериментальное) = −2,16,

t (критическое) = −1,671.

Мы видим, что статистика критерия попадает в критическую область. Если бы [illegible] μ_2) была истинной, такое значение выборочной статистики [illegible] маловероятным, поэтому мы можем отклонить нулевую гипотезу. Имеет место статистически значимое различие [illegible] различие настолько велико, что вероятность его возникновения вследствие простой случайности крайне мала) в [illegible] проживающих в центре города, и семей, проживающих в пригородах. Более того, семьи, проживающие в центре города, значительно больше в размере. Статистика критерия и распределение выборочных статистик изображены на рис. 9.1. *В качестве практического задания, касающегося проверки значимости различий между средними значениями выборок при малых размерах выборок, см. задачи 9.7–9.9.*

Рис. 9.1

Распределение выборочных статистик с изображенной критической областью и [illegible]

Практический пример 9.1

Анкеты со шкалой, измерявшей удовлетворенность семейной жизнью, были розданы лицам, входившим в выборку, состоящую [illegible]

[illegible] нулевой гипотезы (H_0: $\mu_1 = \mu_2$) [illegible]

Оценка стандартного отклонения распределения выборочных статистик будет [illegible]

[illegible]

Z (экспериментальное) = [illegible]

[illegible]

Если альфа имеет значение 0,05, тогда критическая область будет начинаться с Z (критическое) = ±1,96. Если Z (экспериментальное) равно 5,56, мы можем отклонить нулевую гипотезу. Эта проверка подтверждает то заключение, что [illegible]

9.4 Проверка гипотез о выборочных пропорциях (большие выборки)

[illegible] пропорциями средними. Нулевая гипотеза утверждает [illegible] сравненной. Выборочные пропорции служат [illegible] критерия, вычисляемой на шаге 4, которая затем [illegible] значением. При больших размерах выборок (если сумма [illegible] 100) для нахождения границы критической области [illegible] Z [illegible] ение. В этой книге мы не будем рассматривать про[illegible]

[illegible] шаге 6 представлена в виде формулы 9.7. [illegible] для s напоминает нам о том, что теперь мы имеем дело с распределением разностей пропорций выборок.

Формула 9.7 Z (экспериментальное) = [illegible]

где $(P_{s1} - P_{s2})$ = разность между пропорциями выборки

$(P_{u1} - P_{u2})$ = разность между пропорциями генеральной совокупности

σ_{p-p} = стандартное отклонение распределения разностей выборочных [illegible]

Как в случае с одной выборкой, предполагается, что, исходя из нулевой гипотезы, второй член в числителе равен 0. Поэтому формула сокращается.

Формула 9.8 Z (экспериментальное) = [illegible]

Значение стандартного отклонения распределения выборочных статистик различий в пропорциях выборок вычисляется с помощью формулы 9.9.

Формула 9.9 [illegible]

Использование этой формулы требует введения оценки для значения P_u. Это можно сделать путем объединения информации из двух выборок, как показано в формуле 9.10.

Формула 9.10 [illegible]

[illegible] Z (экспериментальное), мы должны начать с формулы [illegible] 9.8. [illegible] мер, чтобы сделать эти процедуры более понятными.

[illegible] чернокожих и белых пожилых граж[illegible] участию в добровольных организациях [illegible] группы, отнесенные к классу "много", и размеры выборок [illegible]

Выборка 1 — Чернокожие пожилые граж[illegible]

Выборка 2 — [illegible]

Практический пример 9 2

Отличается ли отношение к добрачным сексуальным связям для различного [illegible] был задан вопрос [illegible] [illegible] [illegible] которые считают добрачный секс вполне неправильным составляют

[illegible]	Выборка 2
[illegible]	Мужчины
	$P = [illegible]$
	$N = 4[illegible]7$

Это вся информация, в которой мы нуждаемся, чтобы проверить нулевую гипотезу (H_0: $P_1 = P_2$). Оценка для пропорции в генеральной совокупности равна

$$P_u = \frac{N_1 P_{s1} + N_2 P_{s2}}{N_1 + N_2}$$

[illegible]

$$P_u = \frac{290{,}94}{867}$$

[illegible]

Оценка стандартного отклонения распределения выборочных статистик будет равна:

$$\sigma_{p-p} = \sqrt{P_u(1 - P_u)}\sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

$$\sigma_{p-p} = \sqrt{(0{,}34)(0{,}66)}\sqrt{\frac{450+417}{(450)(417)}}$$

$$\sigma_{p-p} = \sqrt{0{,}2244}\sqrt{\frac{867}{[illegible]}}$$

[illegible]

$$Z\text{ (экспериментальное)} = \frac{0{,}35 - 0{,}32}{0{,}032}$$

$$Z\text{ (экспериментальное)} = \frac{0{,}030}{[illegible]}$$

$$Z\text{ (экспериментальное)} = [illegible]$$

Если альфа имеет значение 0,05, тогда критическая область будет начинаться с Z (критическое) $= \pm 1{,}96$. Если Z (экспериментальное) равно 0,94, мы не можем отклонить нулевую гипотезу. Эта проверка подтверждает то заключение, что не существует статистически значимого различия в отношении мужчин и женщин к добрачным сексуальным отношениям.

Шаг 1. Принятие предположений

Модель: Независимые случайные выборки
Номинальная шкала
Нормальное распределение выборочных статистик

Шаг 2. Формулирование нулевой гипотезы

Так как не было предположений о направлении, будем использовать двусторонний критерий [illegible]

$$H_0: P_1 = P_2$$

$$(H_1: P_1 \neq P_2)$$

Шаг 3. Подбор распределения выборочных статистик и установление критической области

Так как выборки имеют большой размер, [illegible] Установив для альфа значение 0,05, имеем:

Распределение выборочных статистик $= Z$ [illegible]
альфа = 0,05, двусторонний
Z (критическое) $= \pm 1{,}96$

Шаг 4. Вычисление статистики критерия

Помните, что следует [illegible] затем подставить результирующее значение в формулу для вычисления [illegible] и найти Z (экспериментальное) по формуле [illegible]

$$P_u = \frac{N_1 P_{s1} + N_2 P_{s2}}{N_1 + N_2}$$

$$P_u = \frac{[illegible]}{83 + 103}$$

$$[illegible] = 0{,}29$$

$$\sigma = \sqrt{[illegible]}\sqrt{\frac{N_1 + N_2}{[illegible]}}$$

$$\sigma = \sqrt{(0{,}29)(0{,}71)}\sqrt{\frac{(83+103)}{(83)(103)}};$$

$$\sigma = (0{,}45)([illegible])$$

$$\sigma = (0{,}07)$$

$$Z(\text{экспериментальное}) = \frac{(P_1 - P_2)}{\sigma};$$

$$Z(\text{экспериментальное}) = \frac{0{,}34 - 0{,}25}{0{,}07};$$

$$Z(\text{экспериментальное}) = 1{,}29$$

Шаг 5. Принятие решения

Так как статистика критерия, Z [illegible] 1,29 [illegible] тическую область, указанную через Z (критическое) = [illegible] нить нулевую гипотезу. Ве[illegible] различий между [illegible] ожидаемых различий [illegible]

[illegible]

[illegible] в научных [illegible] указывают только [illegible]

[illegible] значение альфа [illegible] степеней свободы [illegible] и размер выбор[illegible]. Результаты [illegible] в профессиональной литературе [illegible] между средними зна[illegible] города) и определены, как значимые (t = –2,16, число степеней свободы = 77, $p < 0.05$)".

Обратите внимание на то, что значение альфа было указано как "$p < 0.05$". Такая запись является сокращением от "вероятность (probability) того, что различие такой величины возникло вследствие простой случайности (при условии, что нулевая гипотеза об отсутствии различий истинна), меньше 0,05" и хорошей иллюстрацией того, как исследователь может передать большой объем информации в нескольких символах. Подобным образом наша длинная фраза "статистика критерия попадает в критическую область, и поэтому нулевую гипотезу следует отклонить" была заменена на более короткую и простую: "различия... были... определены как значимые".

Когда исследователи нуждаются в том, чтобы указать результаты нескольких проверок значимости, они часто используют сводную таблицу для указания информации, относящейся к выборке, а также для указания того, являются ли различия значимыми при данном значении альфа. Если вы читаете описание и анализ таких таблиц, проведенные исследователем, у вас, должно быть, возникают определенные сложности в их интерпретации и понимании. В конце концов, эти комментарии о том, каким образом отчеты о проверке значимости составляются в профессиональной литературе, применимы и для проверки гипотез, описанной в части II этого материала.

Статистика в профессиональной литературе

Социологи [illegible] ной выборки населения США для изучения подходов к воспитанию в консервативных протестантских семьях. Они обратили внимание на то, что существует два представле[illegible] исследователи и большая часть общественности, является представление о том, что [illegible] ния (часто физические), строгую дисциплину и деспотичное утверждение власти. В [illegible] привязанность к своим детям и вряд ли когда-нибудь словесно унизят или испугают [illegible]

[illegible] ное поведение и плохую успеваемость [illegible] рененное представление о консервативных [illegible] должно быть распространенным.

[illegible]

[illegible] между консервативными и неконсервативны[illegible] представлены в таблице, где проведено [illegible] [illegible] семьями с детьми [illegible] возраста. Авторы утверждают, что эти результаты подтверждают второе, менее распространенное представление о консервативных протестантских семьях.

Различия средних значений частоты повышения голоса родителями, связанные с религиозными различиями

	Дети дошкольного возраста	Дети школьного возраста
[illegible]	$\bar{X}$ [illegible]	$\bar{X}$ [illegible]
	s [illegible]	s [illegible]
[illegible]	$\bar{X}$ [illegible]	$\bar{X}$ [illegible]
	s [illegible]	s [illegible]
	Общее N = [illegible]	[illegible]

Различия в средних значениях частоты повышения голоса родителями, имеющими детей дошкольного возраста (столбец 1), являются значимыми в минимальной степени ($p < 0{,}[illegible]$), так как родители, являющиеся членами консервативных протестантских деноминаций, [illegible] указывают более низкие значения частоты словесной брани. Мы обнаруживаем более устойчивые различия в средних значениях ($p < 0{,}05$) в частоте повышения голоса, указанной родителями, имеющими детей школьного возраста.

[illegible] предположение, что [illegible] протестантских семьях и [illegible] контекстуальные факторы [illegible] могут выявить [illegible] например, физическими наказаниями [illegible]

[illegible]

9.5 Ограничения при проверке гипотез: Статистическая значимость и практическая важность

[illegible] факторы, которые влияют на [illegible] отклонения нулевой гипотезы [illegible]

1. [illegible]
2. [illegible]
3. [illegible]
4. [illegible]

[illegible]

[illegible] является функцией от процедур проверки (т.е. от того, каким образом [illegible] переменных)[illegible] действительность, которую мы и пытаемся исследовать.

Имеет место прямая связь между значением альфа и вероятностью отклонения [illegible] выше процентный показатель всех возможных значений выборочных статистик, попадающих в критическую область, и тем выше вероятность отклонения нулевой гипотезы.

Таким образом, при значении альфа 0,05 отклонить H_0 проще, чем при значении альфа 0,01, а при значении альфа 0,10 это еще легче сделать. Здесь, конечно же, кроется опасность, так как увеличение значения альфа ведет к более частому появлению ошибок первого рода, и может оказаться, что мы объявим маленькие различия статистически значимыми. Подобным образом применение одностороннего критерия повышает вероятность отклонения (при условии, что было указано верное направление).

Последним фактором является размер выборки: чем больше выборка (если все остальные факторы остаются без изменений), тем выше вероятность отклонения H_0. Эта связь может показаться неожиданной, но ее можно понять, если кратко просмотреть формулы, используемые на шаге 4 для вычисления статистики критерия. Во всех этих формулах размер выборки (N) находится в знаменателе [illegible] это эквивалентно тому, как если бы он находился в числителе формулы. [illegible] означает, что значение статистики критерия прямо пропорционально N и что при увеличении одного значения другое также будет увеличиваться. Чтобы продемонстрировать это, рассмотрим таблицу 9.[illegible], в которой приведены значения статистики критерия, соответствующие средним значениям при наблюдениях с одной выборкой, полученным на основании выборок различных размеров. Значение статистики критерия, Z (экспериментальное), растет по мере того, как увеличивается N, хотя значения остальных членов не изменились. Эта связь между высокой вероятностью отклонения H_0 и большими размерами выборки сохраняется для всех проверок значимости.

Таблица 9.1. Статистики критерия для средних значений в случае одной выборки, вычисленные на основании выборок различного размера ($\bar{X}$ = 80, μ = [illegible]9, s = 5)

Размер выборки	Статистики критерия, Z (экспериментальное)
[illegible]	[illegible]
200	2,[illegible]
500	[illegible]

С одной стороны, взаимосвязь между размером выборки и вероятностью [illegible] большие [illegible] является лучшим [illegible] она представляет. Таким образом, к решениям, [illegible] размера, можно относиться с большим доверием, [illegible] выборках малого размера.

С другой стороны, эта взаимосвязь ясно подчеркивает то, что, возможно, явля- [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] быть статистически значимыми. Конечно же, даже при небольших размерах выборки различия, которые в целом являются банальными и не представляющими интереса, могут быть статистически значимыми. Основная мысль заключается в том, что статистическая значимость и теоретическая или практическая важность — не одно и то же. Различие, которое не является статистически значимым, почти никогда не является важным. Однако значимость сама по себе не обеспечивает важности. Даже в тех случаях, когда совершенно ясно, что результаты исследования не возникли вследствие простой случайности, исследователю, кроме того, все еще нужно оценить их важность. Насколько сильно они поддерживают теорию или гипотезу? Насколько четко они согласуются с прогнозом или анализом? Насколько точно они определяют линию действия при решении некоторой задачи? Такие вопросы исследователь должен задать при оценке важности результатов статистической проверки.

Кроме того, следует отметить, что исследователи имеют доступ к некоторым очень мощным методам анализа важности (а не статистической значимости) результатов исследования. Эти статистические показатели, к которым относятся парные меры тесноты связи и многомерные статистические методы, будут приведены в части III и части IV этой книги.

9.6. Интерпретация статистик: существуют ли значимые различия между доходами мужчин и доходами женщин?

В Соединенных Штатах, как и во многих [illegible] странах мира, были предприняты [illegible] чтобы уравнять условия труда для мужчин и для женщин [illegible] Предолжают ли существовать [illegible] [illegible]

[illegible] [illegible] (необходимой информ[illegible] США, но его записи конфиденциальны). [illegible] [illegible] генеральной совокупности между [illegible] различия в среднем размере заработка. В [illegible] будет использоваться общенацио[illegible] (см. приложение Ж), который про[illegible] [illegible] отобранных случайным образом.

[illegible]

Мужчины	[illegible]
[illegible]	$\bar{X}_2$ [illegible]
[illegible]	$s_2 = 18924$
[illegible]	$N_2 = 666$

Возникает впечатление [illegible] гипотеза в данном случае утверждает [illegible] средний доход ($\mu_1 = \mu_2$). Мы пропустим [illegible] по формуле 9.2 (шаг 4), равно 8.78, что [illegible] принятым стандартизованное значение Z (крит[illegible] со значением альфа 0.05. Это различие в средних [illegible] так что мы можем отклонить нулевую гипотезу [illegible] (с вероятностью ошибки 5%) в том, что средние значения [illegible] тывают значительно больше, чем женщины.

Является ли расхождение в [illegible] Если бы женщины были [illegible]

ни, это можно было бы считать одной из причин различий в доходах. Используя еще раз данные общенационального социологического обследования 1996 года для получения информации о количестве лет образования для случайных выборок мужчин и женщин, мы получили следующую информацию:

[illegible]

[illegible] образования (на один год больше, чем среднее образование) и что для мужчин среднее значение выше, чем для женщин. Разница выглядит довольно небольшой (по крайней мере, в сравнении с указанной выше разницей в доходах), и, без сомнения [illegible] является основанием для отклонения нулевой гипотезы при значении альфа 0.05. Таким образом, мужчины и женщины имеют, по существу, равные уровни образования, и этот фактор нельзя считать причиной различий в доходах.

Эти результаты предполагают, что женщины получают от своего образования меньший результат в смысле дохода. Хотя они практически равны с мужчинами в вопросе подготовки к профессиональной деятельности (по крайней мере, с точки зрения уровня образования), сохраняются значительные различия в доходах. Может ли быть, что различие доходов частично объясняется тем, что мужчины и женщины выбирают разные профессии? Часто утверждают, что мужчины более склонны к доходным престижным профессиям (юрист, доктор), тогда как женщины больше инте- [illegible]

Мы можем проверить эту идею, [illegible] до 300 профессиям. По этим данным [illegible] я какие преимущественно [illegible] профессий, была составлена [illegible] выборки. Нулевая гипо- [illegible]

[illegible]

[illegible] описательные статистические показатели для обеих групп

Преимущественно мужские профессии	[illegible]
[illegible]	[illegible]
[illegible]	$s_2 = 179.70$
[illegible]	[illegible]

Мы [illegible] плате составляет приблизительно $ 32 в неделю или $7000 в год. Z (экспериментальное), вычисленное на основании этих статистических показателей, равно 3,62. Это выше, чем обычное Z (критическое), равное ±1.96. Мы можем отклонить нулевую [illegible] высокий еженедельный заработок, чем преимущественно женским профессиям. Эти результаты свидетельствуют в пользу идеи о том, что расхождения между полами в вопросе доходов частично вызваны тем фактом, что женщинам более свойственно выбирать профессии с более низким заработком. Эти различия в доходах будут существовать до тех пор, пока между представителями разного пола не будет более равномерного разделения профессий или пока заработок не будет уравновешен ка[illegible]

Заключение

1. При проведении исследований часто встречается ситуация, когда нужно прове- [illegible] ные статистики вычисляются для разных выборок из каждой генеральной совокупности, после чего выполняется проверка значимости различий между выборками, которая является методом логического вывода о различиях между данными генеральными совокупностями.
2. Когда информация о выборках выражается в виде выборочных средних и [illegible] ние N высоко, тогда для задания критической области применяется Z-распределение. Когда значение N невелико, для задания критической области использу- [illegible]
3. Разность выборочных пропорций также можно проверить на значимость. При [illegible] области применяется Z-распределение.
4. При проверке гипотез на вероятность отклонения нулевой гипотезы влияет множество факторов: величина различий, [illegible] и размер выборки. [illegible] рального или двустороннего критерия и размер выборки. [illegible] мость — не то же самое, что теоретическая или практическая [illegible] различия были выявлены, как статистически значимые, [illegible] жен показать существенность или важность [illegible] тические показатели, представленные в [illegible]

...рументами, которые нужны нам для того, чтобы заниматься другими вопросами, помимо статистической значимости.

Сводка формул

Статистика критерия для средних значений в случае двух выборок, большие размеры выборок

9.1 $$Z\text{(экспериментальное)} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}-\bar{x}}}$$

Статистика критерия для средних значений в случае двух выборок, большие размеры выборок (упрощенная формула)

9.2 $$Z\text{(экспериментальное)} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{x}-\bar{x}}}$$

Стандартное отклонение распределения разности средних значений выборок, большие размеры выборок

9.3 $$\sigma_{\bar{x}-\bar{x}} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

Обобщенная оценка стандартного отклонения распределения разности средних значений выборок, большие размеры выборок

9.4 $$\sigma_{\bar{x}-\bar{x}} = \sqrt{\frac{s_1^2}{N_2 - 1} + \frac{s_2^2}{N_1 - 1}}$$

Обобщенная оценка стандартного отклонения распределения разности средних значений выборок, небольшие размеры выборок

9.5 $$\sigma_{\bar{x}-\bar{x}} = \sqrt{\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}}\sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

Статистика критерия для средних значений в случае двух выборок, небольшие размеры выборок

9.6 $$t\text{(экспериментальное)} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{x}-\bar{x}}}$$

Статистика критерия для пропорций двух выборок, большие размеры выборок

9.7 $$Z\text{(экспериментальное)} = \frac{(P_{s1} - P_{s2}) - (P_{u1} - P_{u2})}{\sigma_{p-p}}$$

Статистика критерия для пропорций двух выборок, большие размеры выборок (упрощенная формула)

9.8 $$Z\text{(экспериментальное)} = \frac{(P_{s1} - P_{s2})}{\sigma_{p-p}}$$

Стандартное отклонение распределения разностей выборочных пропорций, большие размеры выборок

9.9 $$\sigma_{p-p} = \sqrt{P_u(1 - P_u)}\sqrt{(N_1 + N_2)/N_1 N_2}$$

Обобщенная оценка пропорции генеральной совокупности, большие размеры выборок

9.10 $$P_u = \frac{N_1 P_{s1} + N_2 P_{s2}}{N_1 + N_2}$$

Словарь

Независимые случайные выборки. Случайные выборки, сформированные таким образом, при котором выбор каждого отдельного объекта для одной выборки не оказывает влияния на вероятность выбора каждого отдельного объекта для других выборок.

Обобщенная оценка. Оценка стандартного отклонения распределения разностей выборочных средних, основанная на стандартных отклонениях обеих выборок.

σ_{p-p}. Символ стандартного отклонения распределения разностей выборочных пропорций.

$\sigma_{\bar{x}-\bar{x}}$. Символ стандартного отклонения распределения разностей средних значений выборок.

Мультимедийные ресурсы

The Wadsworth Sociology Resource Center: Virtual Society (Виртуальное сообщество Центра социологических ресурсов Вадсворта)

[illegible]

Для доступа к широкому диапазону студенческих ресурсов посетите web-сайт, сопровождающий настоящее издание книги *Statistics: A Tool for Social Research*. Щелкните на разделе "Студенческие ресурсы" (Student Resources) web-сайта книги, чтобы получить доступ к следующим инструментам исследования.

- Обзор основ математики,
- Обучающие карточки,
- Дополнительные задачи по главам
- Обзор статистики
- Интернет-ссылки,
- Таблица случайных чисел
- Примеры и упражнения с использованием статистических программ MicroCase и SPSS;
- Функциональные диаграммы "[illegible]"
- Проверка гипотез для [illegible]

Задачи

9.1 [illegible] статистических показателей выборок, используя пятишаговый алгоритм [illegible]. *Подсказка. Помните, что прежде чем начинать вычисление значения по формуле 9.2, следует вычислить значение формулы 9.4. Кроме того, в формуле 9.4 следует выполнять математические операции в правильной последовательности. Во-первых, следует вычислить квадраты среднеквадратических отклонений каждой выборки, затем нужно разделить это число на N − 1, сложить полученные значения и найти квадратный корень суммы.*

а

Выборка 1	Выборка 2
$\bar{X} = 72{,}5$	$\bar{X}_2 = 76{,}0$
$s_1 = $ [illegible]	$s_2 = 10{,}2$
$N_1 = 136$	$N_2 = 257$

б

Выборка 1	Выборка 2
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

9.2. SOC

Компания Gessner and Healey составила анкеты для заполнения студентами. Помимо прочего, анкеты содержали шкалу, предназначенную для измерения отношения к межличностному насилию (где более высокие значения указы[illegible]. Ниже приведены результаты, полученные с разделением по половому, расовому и социальному признаку.

а

Мужчины	Женщины
[illegible]	$\bar{X}_2 = 2{,}79$
[illegible]	[illegible]
[illegible]	[illegible]

б

Черные	Белые
[illegible]	[illegible]
[illegible]	[illegible]
$N = 4$[illegible]	[illegible]

в

Белые воротнички	Синие воротнички
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

г [illegible]

9.3. SOC

[illegible] для случайных выборок из баскетболистов и футболистов. Является ли наблюдаемое различие значимым? Напишите два-три предложения, объясняющих различие.

Баскетболисты	Футболисты
$\bar{X}_1 = 460$	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

Что можно сказать [illegible]?

Мужчины	Женщины
[illegible] 452	[illegible]
[illegible] 88	[illegible]
[illegible]	[illegible]

9.4. PA

Несколько лет назад пожарное отделение города [illegible], Канзас, начало нанимать представителей групп меньшинств через Программу позитивных действий. С точки зрения оценок эффективности, собранных начальниками, как [illegible] работников, [illegible] по Программе позитивных действий?

Программа позитивных действий	Обычная процедура
[illegible] 5,2	[illegible]
[illegible]	[illegible]
$N_1 = 97$	$N_2 = 100$

Напишите одно-два предложения с объяснением полученных результатов.

9.5. SOC

Можно ли сказать, что семьи, относящиеся к среднему классу, поддерживают более тесные связи с родственниками, чем семьи, относящиеся к рабочему классу? Напишите абзац, в котором подведите итоги по результатам этих проверок.

а [illegible] они наносят в среднем 7,3 визита близким родственникам [illegible] семьи-представители рабочего класса [illegible] в среднем 8,2 визита близким родственникам [illegible].

Средний класс	Рабочий класс
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

[illegible] почты и месяц близким [illegible] класса совершают в сред[illegible] сообщений электронной почты [illegible]

Телефонные звонки	
Средний класс	Рабочий класс
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

Сообщения электронной почты	
Средний класс	Рабочий класс
$\bar{X}_1 = 8.7$	$\bar{X}_2 = 5.7$
$s_1 = 0.5$	$s_2 = 1.1$
$N_1 = $ [illegible]	$N_2 = $ [illegible]

9.6. SOC

[illegible] живущие в общежитиях, больше вовлечены [illegible] чем те, кто приезжают в студенческий городок [illegible]? [illegible] данные о среднем количестве часов в неделю [illegible] внеурочной деятельности. Являются ли [illegible] случайными выборками студентов, приезжающих [illegible] живущих в общежитиях.

Жители общежитий	Приезжающие
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

9.7. SOC

Можно ли сказать, что пожилые жители, живущие в домах престарелых, более [illegible] возрастов? Напишите одно-два предложения, где объясните полученные результаты. *Подсказка: помните о том, какие формулы следует использовать для выборок небольших размеров.*

а. [illegible] домах престарелых [illegible] разговоров с соседями [illegible] случайной выборки людей, живущих в обществе [illegible] возрастов, сообщили о том, что для них этот показатель [illegible] значимым.

Дом престарелых	Общество людей всех возрастов
$\bar{X}_1 = $ [illegible]	$\bar{X}_2 = $ [illegible]
$s_1 = $ [illegible]	$s_2 = $ [illegible]
$N_1 = 43$	$N_2 = 37$

9.8. SW

Как директор местного филиала Boys Club вы на протяжении многих лет [illegible] заявляли, что членство в вашем клубе снижает уровень подростковой преступности. Теперь же один циничный член агентства [illegible] [illegible] клуба. Респондентам в этих выборках были заданы вопросы относительно их вовлеченности в преступную деятельность. Каждого респондента попросили [illegible] прошедшего года. Полученные результаты приведены ниже (среднее число фактов правонарушения, признанных респондентами). Что вы можете сказать агентству, финансирующему вашу организацию?

Члены	Нечлены
$\bar{X}_1 = 0.3$	$\bar{X}_2 = 2.3$
$s_1 = 2$ [illegible]	$s_2 = $ [illegible]
$N_1 = 40$	$N_2 = 55$

9.9. SOC

Можно ли сказать, что профессиональные футбольные команды в настоящее время больше практикуют наступательную тактику? Зарабатывают ли они в среднем больше очков, чем команды в былые годы? Игры были выбраны случайным образом из последнего сезона и из сезона, прошедшего 10 лет назад. Является ли разница между показателями среднего количества очков за игру значительной?

Последний сезон	10 лет назад
$\bar{X}_1 = $ [illegible]	$\bar{X}_2 = $ [illegible]
$s_1 = $ [illegible]	$s_2 = $ [illegible]
$N_1 = 35$	$N_2 = $ [illegible]

9.10. Для каждой задачи выполните проверку значимости различий. *Подсказка: при проверке значимости различий пропорций помните, что следует вычислить формулу 9.10, затем использовать формулу 9.9, а затем 9.8.*

а.

Выборка 1	Выборка 2
$P_{s1} = $ [illegible]	$P_{s2} = 0.$[illegible]
$N_1 = $ [illegible]	$N_2 = $ [illegible]

б.

Выборка 1	Выборка 2
$P_{s1} = 0.62$	$P_{s2} = 0.60$
$N_1 = 532$	$N_2 = 479$

9.11. CJ

Приблизительно половина [illegible] прошли специальный курс следственного производства [illegible] преступлений [illegible]

[illegible] выборки офицеров. Не про- [illegible]

Прослушавшие курс	Не прослушавшие курс
[illegible]	[illegible]
[illegible]	[illegible]

9.12 [illegible]

[illegible] оценить несколько эксперимен- [illegible] в котором подведите итоги по результа- [illegible] новые программы дали положитель- [illegible]

а. [illegible] для консультирования при разводах. Глав- [illegible] программы — ее консультанты, которыми являются семей- [illegible] командах. Около половины всех клиентов были [illegible] участвовать в этой программе, тогда как дру- [illegible] в обычной программе, и для обеих групп были [illegible] пропорции случаев, которые в конечном итоге закончились раз- [illegible] Ниже представлены результаты для случайных выборок, полученных [illegible] участников обеих программ. Если смотреть с точки зрения предот- [illegible]

Специальная программа	Обычная программа
[illegible]	[illegible]
[illegible]	[illegible]

б. Агентство также проводит эксперименты с равноправным воспитанием для детей, растущих в неполных семьях. Около половины всех клиентов были назначены на участие в равноправном воспитании. По прошествии года су- [illegible] случайной выборки детей, участвовавших в новой программе, со случайной выборкой детей, не получивших равноправного воспитания. Можно ли сказать, что програм- ма дала по[illegible]тельные результаты, с точки зрения процентного показате- ля детей, которые были отнесены к категории очень изменился?

Равноправное воспитание	Отсутствие равноправного воспитания
P_s = 0.10	P = [illegible]
[illegible]	[illegible]

9.13 SOC

[illegible]

ности, подавших заявление на поступление в аспирантуру, пропорция студен- тов-представителей своей специальности, принятых в аспирантуру, и доля тех, кто закончил аспирантуру. На основании полученных данных можно ли ска- зать, что есть значительная разница в качестве программ?

а. Пропорция студентов-представителей своей специальности, подавших за- явление на поступление в аспирантуру

Социология	Психология
P = [illegible]	P = [illegible]
N = [illegible]	N = [illegible]

б. Пропорция студентов-представителей своей специальности, принятых в ас- пирантуру

Социология	Психология
P = [illegible] 75	P = 0.85
N = [illegible]	N = [illegible]

в. Пропорция тех, кто закончил обучение в аспирантуре

Социология	Психология
P = [illegible]	P = 0.69
N = [illegible]	N = [illegible]

9.14 CJ

Местный начальник полиции несколько лет назад начал программу гласнос- ти в борьбе с преступностью и он интересуется, дает ли эта программа поло- жительные результаты. Программа направлена на регулярное информирова- ние о нераскрытых насильственных преступлениях в местных средствах массовой информации и предлагает денежные вознаграждения за информа- цию, приводящую к арестам. Можно ли сказать, что обнародованным преступ- лениям соответствует более высокая вероятность их раскрытия, чем другим на- сильственным преступлениям? Ниже представлены результаты, полученные по обеим выборкам

Обнародованные преступления, раскрытые в меру доказательств, достаточных для проведения ареста	Необнародованные преступления, раскрытые в меру доказательств, достаточных для проведения ареста
P_s = 0.35	P_s = 0.25
N = 178	N_2 = 212

9.15 SOC

Ниже представлены некоторые результаты общенационального [illegible] [illegible] значимости различий.

а. Пропорция [illegible] жи, покупки и хранения оружия (GUNLAW)

Мужчины	Женщины
[illegible]	[illegible]

б. Пропорция согласных с тем, что добрачная половая жизнь — это всегда не [illegible] (PREMARSX)

Мужчины	Женщины
[illegible]	$P_u = 0{,}29$

в. [illegible]

Мужчины	Женщины
[illegible]	[illegible]

г. [illegible]

Мужчины	Женщины
[illegible]	[illegible]

д. [illegible] (ATTEN)

Мужчины	Женщины
[illegible]	[illegible]

е. [illegible]

Мужчины	Женщины
$\bar{X}$ [illegible]	$\bar{X}$ [illegible]
$N = 620$	$N =$ [illegible]

SPSS for Windows

Использование SPSS for Windows для проверки значимости различий между двумя средними значениями

Демонстрационный пример с использованием SPSS 9.1. Кто больше смотрит телевизор – мужчины или женщины?

SPSS for Windows [illegible]

функцию Independent-Samples T Test (T-критерий для независимых [illegible]) [illegible] различий между мужчинами и женщинами [illegible] просмотром телевизора [illegible] различие между средними значениями выборок для мужчин и для женщин, мы можем сделать вывод, что генеральные совокупности (все взрослые мужчины и женщины в США) отличаются по этой переменной.

Запустите *SPSS for Windows* и загрузите базу данных GSS1998. В строке главного [illegible] Analyze (Анализ), затем [illegible] Compare Means ([illegible]ние средних), и затем по пункту Independent-Samples T Test. Откроется аналоговое окно Independent-Samples T Test, в котором [illegible] Найдите в этом списке переменную *tvhours — время, проведенное за просмотром телевизора* (метка переменной — HOURS PER DAY WATCHING TV), наведите на нее курсор и щелкните по стрелке вверх в центре окна, чтобы переместить переменную *tvhours* в список Test Variable(s) (Проверяемая переменная). Затем найдите и выделите переменную *sex* (метка — RESPONDENTS SEX) и щелкните по стрелке вниз, расположенной в середине экрана, чтобы переместить переменную *sex* в окно Grouping Variable (Группирующая переменная). В окне Grouping Variable появятся два вопросительных знака, и кнопка Define Groups станет активной. SPSS необходимо знать, какие наблюдения должны относиться к каким группам. В данном случае инструкции, которые следует дать системе, довольно просты. Мужчины (для которых переменная *sex* имеет значение 1) включаются в группу 1, а женщины (значение 2) — в группу 2.

Щелкните по кнопке Define Groups (Определить группы), после чего появится окно Define Groups. В окне, соответствующем группе 1, появится мигающий курсор — SPSS запрашивает значение, определяющее, какие наблюдения должны включаться в эту группу. Введите в этом окне 1 (т.е. мужчины), после чего щелкните по окну, соответствующему группе 2, и введите 2 (т.е. женщины). Щелкните по кнопке Continue, [illegible] Independent-Samples T Test [illegible] OK [illegible] удален в целях экономии места).

В первом блоке выходных данных представлены показатели описательной статистики. В выборке было 502 мужчины, и они смотрели телевизор в среднем по 2,84 часа в день со стандартным отклонением, равным [illegible] смотрели телевизор в среднем 2,87 часа в день со стандартным [illegible]

Групповая статистика

	Пол респондентов	N	Среднее значение	Стандартное отклонение	Ошибка среднего
[illegible]	Мужчины	[illegible]	[illegible]84	[illegible]	[illegible]
[illegible]	Женщины	[illegible]	[illegible]	[illegible]	[illegible]

Проверка независимых выборок

		Критерий Левене для проверки равенства дисперсий		Критерий для равенства средних				
		F	Знач.	t	Ст. св.	Знач. (2-сторонняя)	Разность между средними	Разность между стандартными отклонениями
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	,822	[illegible]	[illegible]
при мнимом значении	Не предполагается							
				[illegible]	[illegible]	,822	[illegible]	[illegible]

Мы опустим первые столбцы следующего блока выходных данных, в котором приводятся результаты проверки равенства дисперсий генеральных совокупностей. В этом блоке представлены результаты проверки на значимость. *SPSS для Windows* проводит отдельную проверку для каждого предположения относительно дисперсии генеральной совокупности (см. разделы 9.2 и 9.3), но мы просмотрим только раздел Equal variances assumed (Предполагается равенство дисперсий), представленный в первой строке. Это в целом соответствует той модели, которую мы использовали в этой главе.

SPSS for Windows выводит значение *t* (-0,225), количество степеней свободы df = ,32 и статистической значимости [illegible] ... которое является нашим визуальным индикатором значимости. Мы [illegible] ...

Демонстрационный пример с использованием SPSS 9.2. Использование команды COMPUTE для проверки различий отношения представителей разных полов к абортам

[illegible] ... команду **Compute** для со[illegible] ... ия отношения к абортам путем [illegible] ... *abhlth* и *abany*). Помните, что пос[illegible] ...

[illegible] переменную, следуя инструкциям, приведенным в демонстрационном приме[illegible] ...

[illegible] ... Independent-Samples T Test [illegible] ... монстрационном примере 9., и переместите в окно **Test Variable(s)** переменную [illegible] ... ру назначения переменной *sex* в качестве переменной группирования. Ваш [illegible] ...

Групповая статистика

	Пол респондентов	N	Среднее значение	Стандартное отклонение	Ошибка среднего
ABINDEX	Мужчины	[illegible]	[illegible]	[illegible]	[illegible]
	Женщины	[illegible]	[illegible]	[illegible]	[illegible]

Проверка независимых выборок

		Критерий Левене для проверки равенства дисперсий		t-критерий для равенства средних		
		F	Знач.	t	Ст. св.	Знач. (2-стор.)
ABINDEX	Предполагается равенство дисперсий	2,22	,37	1,997	853	0,046
	Не предполагается равенство дисперсий			[illegible]	[illegible]	

Разность между средними	Разность между стандартными отклонениями	95% доверительный интервал разности: Нижняя граница	Верхняя граница
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Выборочные средние значения выборок [illegible] ... с 2,7275), однако, статистика критерия [illegible] ... чение Знач. (2-сторонний) равно [illegible] ... знак значимости результата). [illegible] ...

противниками абортов, чем мужчины (по крайней мере с точки зрения причин, которые выражают переменные *abany* и *abdefd*).

Наконец, позвольте мне обратить ваше внимание на пункт Number of Cases (Ко [illegible] из выборки [illegible] GSS [illegible] 1998 [illegible] чем на 1500 человек. Что же случилось со всеми этими наблюдениями?

Как вы помните из приложения Б, многие вопросы GSS были представлены только двум третям выборки. Более того, как я говорил в конце главы 4, команда Compute исключает наблюдение из вычислений, если в нем отсутствует значение по какой-либо из составляющих переменных. Таким образом, для большинства отсутствующих наблюдений эти вопросы не были заданы, а остальные не ответили на один или на оба вопроса, являющихся составляющими вопроса об отношении к абортам. Это явление сокращения размера выборки является довольно распространенной проблемой в социологических исследованиях, и оно в некоторой степени может подвергнуть опасности целостность исследования.

Упражнения

9.1. Существуют ли значимые различия между мужчинами и женщинами в таких вопросах, как престижность профессии, образование и доход? Используя демонстрационный пример 9.1 в качестве примера, подставьте *prestg80*, *educ* и [illegible] итоги по результатам этой проверки [illegible] некоторые идеи, касающиеся написания такого отчета, можно почерпнуть из раздела [illegible]

9.2. Используя демонстрационный пример 9.2 в качестве примера, постройте обобщенную шкалу отношения к традиционным ролям полов, подставив *fehelp* и [illegible] *fepol* [illegible] зуя переменную *sex* в качестве независимой переменной. Напишите одно-два предложения, в которых поясните итоги по результатам этой проверки.

9.3. [illegible] переменных можно объяснить различия во мнениях относительно абортов. Выберите [illegible] переменных, различных по переменной [illegible] проверьте значимость различий [illegible] независимая переменная [illegible] зависимая переменная имела только две категории, поэтому ваши возможности выбора в базе данных 1998 GSS довольно ограничены. Один из вариантов в данном случае [illegible] переменных, которые содержат более двух категорий, с помощью команды Recode SPSS [illegible] Collapse Categories [illegible] Используя демонстрационный пример 9.2 [illegible] выбранных вами независимых переменных [illegible] SPSS [illegible] одно-два предложения [illegible] этих проверок.

10 Проверка статистических гипотез III: Дисперсионный анализ

Цели изучения Изучив данную главу, вы сможете

1. Идентифицировать и приводить примеры ситуаций, в которых уместно применение дисперсионного анализа.
2. Объяснить логику процесса проверки статистических гипотез применительно к дисперсионному анализу.
3. Применять дисперсионный анализ, руководствуясь пятишаговым алгоритмом процесса принятия решений, правильно интерпретировать полученные результаты.
4. Дать определение и пояснить смысл таких понятий, как дисперсия генеральной совокупности, общая сумма квадратов, межгрупповая сумма квадратов, внутригрупповая сумма квадратов, оценки среднего квадрата, апостериорные критерии (post hoc tests).
5. Объяснить, в чем отличие между статистической значимостью и важностью зависимости между переменными.

10.1. Введение

В этой главе мы рассмотрим очень гибкий и широко используемую проверку значимости различий, называемую **дисперсионным анализом** (analysis of variance — ANOVA, сокращенно — ДА). Этот метод можно использовать в ряде ситуаций, когда рассмотренные нами ранее критерии оказываются менее оптимальными или вообще не применимыми. ДА был разработан с целью использования в случае зависимых переменных, измеряемых по интервальной шкале или шкале отношений, и является мощным инструментом анализа наиболее сложных и точных измерений, с которыми нам когда-либо придется столкнуться.

[illegible] ДА как [illegible] (критерия Стьюдента), посредством которого проверяется [illegible] между средними двух выборок (t-критерий мы рассмотрели в главе [illegible] критерий применим только в ситуации, когда [illegible] только две категории (например, *протестанты* и *католики* [illegible] применим для независимых переменных, имеющих [illegible] пример, *протестанты, католики, иудеи, [illegible]*

[illegible] ременного?" Ведь мнение по поводу смертной казни относится к морали и должно отражать религиозные взгляды человека.

Пусть мы ввели интервальную шкалу или шкалу отношений, по которой можно измерить, насколько активно опрашиваемый поддерживает смертную казнь. Провели опрос для случайной выборки, включающей протестантов, католиков, иудеев, атеистов. Итак, мы имеем 4 категории субъектов и хотим выяснить, изменяется ли значимым образом конкретное отношение или мнение в зависимости от категории (в нашем случае — религии), к которой относится данный человек. Мы также хотим ответить и на другие вопросы, например, "Какие религии характеризуются наименьшей (наибольшей) поддержкой смертной казни?", "Является ли степень поддержки протестантами смертной казни значимо большей по сравнению с католиками или иудеями?", "Как атеисты ведут себя по сравнению с верующими?" Дисперсионный анализ предоставляет очень полезный статистический контекст, в котором можно рассмотреть перечисленные вопросы.

10.2. Логика дисперсионного анализа

[illegible] $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (отли[illegible] — расширенная версия нулевой гипотезы t-критерия для двух выборок) [illegible] казни связана с религией.

Если нулевая гипотеза [illegible] носями справедлива [illegible] что средние вряд ли будут в точ[illegible], когда нулевая гипотеза на самом деле [illegible] ядется некоторая ошибка [illegible] для того, чтобы [illegible] генеральные совокуп[illegible]

[illegible] Крайних [illegible]

Таблица 10.1. Поддержка смертной казни в зависимости от религии (данные вымышленные)

	Протестанты	Католики	Иудеи	Атеисты
Среднее	3	[illegible]	[illegible]	[illegible]
Стандартное отклонение	2,4	1,8	2,2	1,7

Таблица 10.2. Поддержка смертной казни в зависимости от религии (данные вымышленные)

	Протестанты	Католики	Иудеи	Атеисты
Среднее	4	[illegible]	[illegible]	[illegible]
Стандартное отклонение	0,5	[illegible]	[illegible]	[illegible]

Давайте теперь рассмотрим второй набор результатов, приведенный в таблице 10.2 (данные вымышленные). Здесь мы видим, что средние баллы для различных категорий значительно различаются между собой, причем для категории *иудеи* наблюдается наименьшая поддержка смертной казни, а для категории *протестанты* — наибольшая. Также мы видим, что стандартные отклонения малы и практически одинаковы по всем категориям, это говорит о том, что внутри каждой категории (религии) изменчивость невелика. Итак, в таблице 10.2 мы видим заметные различия между религиями в сочетании с однородностью внутри каждой религии. Другими словами, есть заметные отличия одной религии от другой, но внутри каждой рели[illegible]

В принципе, ДА и состоит в проведении сравнений типа [illegible] внутри категорий (среди протестантов, среди [illegible] чин между категориями, по отношению к [illegible] вероятность того, что нулевая [illegible] может быть отвергнута. Итак, если [illegible] зависимости от религи[illegible] вполне отличаться [illegible] носительно мал.

10.3. Вычисления при применении ДА

Хотя мы рассматривали ДА как метод для проверки значимости различий между [illegible] дисперсии являются мерами разброса, и что дисперсия является просто возведенным в квадрат стандартным отклонением. Одна оценка дисперсии генеральной совокупности основана на уровне изменчивости внутри каждой категории независимой переменной, другая оценка — на уровне изменчивости между категориями.

Перед тем как строить эти две оценки, нам нужно ввести некоторые новые понятия и статистики. Первое новое понятие — это общая изменчивость значений, которая измеряется величиной, называемой **общей суммой квадратов** или SST (от total sum of squares).

Формула 10.1 $SST = \sum$ [illegible]

Чтобы найти эту величину, нам необходимо взять каждое значение, вычесть из него среднее, возвести полученную разность в квадрат, а затем сложить для получения общей суммы квадратов разностей. Эта формула кажется вам смутно знакомой, ведь это не что иное, как числитель формулы для вычисления дисперсии и стандартного отклонения выборки (см. главу 4). Такое сходство не является неожиданностью, т.к. SST — это тоже способ измерить разброс.

Чтобы построить упомянутые выше две разные оценки дисперсии генеральной совокупности, общую дисперсию (SST) необходимо разделить на две компоненты. [illegible] внутригрупповой суммой квадратов (SSW [illegible]) [illegible] межгрупповой суммой квадратов (SSB [illegible]) [illegible]

Формула 10.2 $SST = SSB + SSW$

Внутригрупповая сумма квадратов [illegible]

Формула 10.3 $SSW = \sum$ [illegible]

где: SSW — внутригрупповая сумма квадратов внутри категорий; [illegible] — [illegible] категории.

Формула [illegible] предписывает нам: взять каждое значение, вычесть из него среднее [illegible], возвести полученную разность в квадрат, а затем [illegible] для каждой категории [illegible] полученные результаты, и мы получим SSW.

[illegible] (SSB) отражает изменчивость между этими [illegible] выступают в роли итоговых статистик для [illegible] среднее по каждой категории выступает в [illegible] SSB имеет вид:

Формула 10.4 $SSB = \sum N_k(\bar{X}_k -$ [illegible]

где: SSB — межгрупповая сумма квадратов;
N_k — количество наблюдений в k-й категории;
$\bar{X}_k$ — среднее k-й категории.

Итак, чтобы найти SSB, рассчитайте общее среднее $\bar{X}$ (как среднее арифметическое средних всех категорий). Затем возьмите среднее для отдельной категории, вычтите из него общее среднее, возведите полученную разность в квадрат, умножьте на количество наблюдений в категории [illegible] и так по всем категориям, а затем просуммируйте.

Давайте остановимся и вспомним, для чего нам нужны SSW и SSB. Если нулевая гипотеза верна, то тогда изменчивость от категории к категории не должна быть большой по сравнению с изменчивостью внутри категорий, т.е. две оценки дисперсии генеральной совокупности — основанные на SSW и SSB, — должны быть примерно равны. Чем больше различие между этими двумя оценками, тем больше вероятность, что нам придется отвергнуть нашу нулевую гипотезу. Если же средние категорий (групп) приблизительно одинаковы, то тогда различия не будут значимыми. Чем больше различие между средними категорий и чем более однородны категории, тем больше вероятность, что различия являются статистически значимыми.

Следующим шагом вычислительной схемы является построение оценок дисперсии генеральной совокупности. Для этого мы разделим каждую сумму квадратов на соответствующее число степеней свободы. Чтобы найти число степеней свободы для SSW, следует из количества наблюдений (N) вычесть количество категорий (k). А чтобы найти число степеней свободы для SSB, следует из количества категорий (k) вычесть единицу. Имеем:

Формула 10.5 $dfw = N - k,$

где: dfw — число степеней свободы для SSW (от degrees of freedom within),
N — количество наблюдений,
k — количество категорий.

Формула 10.6 $dfb = k - 1,$

где: dfb — число степеней свободы для SSB (от degrees of freedom between),
k — количество категорий.

Фактические оценки дисперсии генеральной совокупности, называемые оценками среднего квадрата, рассчитываются [illegible] каждой суммы квадратов на соответствующую ей степень свободы:

Формула 10.7 Внутригрупповой [illegible] $= \frac{SSW}{dfw}$

Формула 10.8 Межгрупповой [illegible] $= \frac{SSB}{dfb}$

[illegible]

Фактическая статистика критерия, используемого в ДА, которую мы рассчитаем [illegible] шагового алгоритма процесса принятия решений, [illegible] ение этой статистики вычисляется по следующей формуле:

$$\text{Формула 10.9} \qquad F = \frac{\text{Межгрупповой средний квадрат}}{\text{Внутригрупповой средний квадрат}}$$

Очевидно, что значение F-отношения (критерия Фишера) будет функцией отношения величины изменчивости между категориями к величине изменчивости внутри категорий. Чем больше изменчивость между категориями по сравнению с изменчивостью внутри категорий, тем больше будет значение F-отношения и тем больше вероятность того, что мы отвергнем нулевую гипотезу.

10.4. Упрощенная схема вычислений

Схема вычислений для ДА, описанная в предыдущем разделе, требует многих шагов и различных формул и на первый взгляд может показаться сложной. Но, как это всегда бывает, как только вы начнете продвигаться от формулы к формуле (конечно, в нужном порядке), то вы увидите, что тут нет ничего сложного, хотя вычисления довольно громоздкие. Поэтому, рискуя злоупотребить вашим терпением, мы все же опишем способ упростить и ускорить вычисления. Для этого придется [illegible] еще несколько формул, которые, в конце концов, приведут к ускорению [illegible] что стоит того!

Формулы, введенные нами до настоящего момента, были формулами-определениями. [illegible] В особенности неудобны для вычислений формулы для SST и SSW, к счастью, для SST есть и удобная для расчета формула.

Формула 10.10 $\quad SST = \sum X^2 - N\bar{X}^2$

[illegible] рассчитать SST по этой формуле, нужно сначала возвести в квадрат значение общего среднего [illegible] значения, убедитесь в том, что вы использовали все значения (наблюдения) [illegible]

После того как вы вычислите значения SST и SSB, можно найти SSW путем преобразования формулы 10.2:

Формула 10.11 $\quad SSW = SST - SSB$

[illegible]

1. [illegible]
2. [illegible]

3. Найдите SSW путем вычитания (по формуле 10.11).
4. Вычислите степени свободы (по формулам 10.5 и 10.6).
5. Постройте две оценки среднего квадрата для дисперсии генеральной совокупности, разделив SSB и SSW соответственно на степени свободы для них (формулы 10.7 и 10.8).
6. Найдите наблюдаемое экспериментальное значение F-отношения, разделив оценку межгруппового среднего квадрата на оценку внутригруппового среднего квадрата (формула 10.9).

Описанные выше вычисления и само применение ДА для проверки значимости различий мы проиллюстрируем в следующих разделах.

10.5. Пример вычислений для ДА

Пусть мы провели опрос по нашей "Шкале поддержки смертной казни" для выборки из 16 человек, равномерно распределенных по четырем религиям (совершенно очевидно, что такая выборка слишком мала для любого серьезного исследования, мы рассматриваем ее только в целях иллюстрации к ДА). Результаты опроса (значения в баллах) приведены в таблице 10.3 вместе с квадратами баллов, средними для каждой категории и общим средним.

Таблица 10.3. Поддержка смертной казни в зависимости от религии, для 16 респондентов (данные вымышленные)

Протестанты X	X^2	Католики X	X^2	Иудеи X	X^2	Атеисты X	X^2
8	64	12	144	12	144	15	225
7	[illegible]	20	400	13	169	16	256
[illegible]	[illegible]	25	625	18	324	23	529
7	[illegible]	27	729	21	441	28	784
51	666	84	1898	64	1078	82	1794
[illegible]		21.0		16.0		20.5	

$\bar{X} = 17.5$

Будем проводить наши вычисления в соответствии с вычислительной схемой, приведенной в конце раздела 10.4. Начнем с нахождения SST по формуле 10.10:

$$SST = \sum X^2 - N\bar{X}^2,$$

$$SST = (666 + 1898 + 1078 + 1794) - (16)(17.5)^2,$$

$$SST = 5436 - (16)(306.25),$$

$$SST = 5436 - 4900,$$

$$SST = 536.$$

[illegible]

$$SSB = \sum N_k(\bar{X}_k - \bar{X})^2$$

$$SSB = (4)(12{,}5 - 17{,}5)^2 + (4)(21\ [illegible]\ 17{,}5)\ [illegible] +$$
$$(4)(16{,}0 - 17{,}5)^2 + (4)(20{,}5\ [illegible]\ 17{,}5)\ [illegible]$$

$$SSB = 4(-5)^2 + 4(3{,}5)^2 + 4(-1{,}5)^2 + 4(3)^2,$$

$$SSB = 100 + 49 + 9 + 36,$$

$$SSB = 194$$

Теперь можно найти SSW путем вычитания (по формуле 10 [illegible])

$$SSW = SST - SSB$$

$$SSW = [illegible]\ 4$$

$$SSW = 342$$

Чтобы найти степени свободы для каждой из сумм квадратов, воспользуемся формулами 10 5 и 10 6 соответственно.

$$dfw = N - k = 16 - 4 = 12$$

$$[illegible]\ k\ [illegible]\ 4\ [illegible] = 3$$

Наконец, мы готовы к построению оценок среднего квадрата для дисперсии [illegible] SSW [illegible] формулу 10.7

$$\text{внутригрупповой средний квадрат} = \frac{SSW}{dfw} = \frac{342}{12} = [illegible]$$

а для получения оценки межгруппового среднего квадрата применим формулу 10.8

$$\text{межгрупповой средний квадрат} = \frac{SSB}{dfb} = \frac{194}{3} = 64{,}67\ .$$

[illegible] находится по формуле 10 9

$$[illegible] = \frac{\text{межгрупповой средний квадрат}}{\text{внутригрупповой средний квадрат}}$$

$$[illegible] = \frac{64{,}67}{[illegible]}$$

$$[illegible] = [illegible]$$

[illegible] следует оценить ее значимость, что мы и произведем [illegible] вычислении этих величин и применении при [illegible]

10 6 Критерий значимости для ДА

[illegible], а также рас[illegible] ДА. Как обычно, мы будем следовать пятишаговой модели — [illegible] способу представления процесса принятия решения)

Шаг 1. Принятие предположений

Модель. Независимые случайные выборки

Зависимая переменная измеряется по интервальной шкале или [illegible]

Генеральные совокупности подчиняются нормальному распределению

Дисперсии генеральных совокупностей равны между собой

Предположения лежащие в основе данной модели довольно жесткие и подчеркивают тот факт, что данный метод следует использовать только для таких зависимых переменных, которые были тщательно и точно измерены. До тех пор, пока объемы выборок равны между собой (или приблизительно равны) ДА может мириться с некоторым нарушением допущений модели, но в ситуациях, когда вы не уверены или имеете выборки, сильно отличающиеся по объему, — советуем вам применить вместо ДА какой-либо альтернативный метод (критерий хи-квадрат, описанный в главе 11 — один из возможных)

Шаг 2 Формулирование нулевой гипотезы

Для ДА нулевая гипотеза всегда состоит в том, что средние генеральных совокупностей, из которых были извлечены наши выборки, равны между собой. Для рассматриваемого примера с четырьмя различными генеральными совокупностями (категориями) нулевая гипотеза будет иметь вид

$$[illegible]$$

где μ_1 представляет собой среднее для категории *[illegible]*, μ_2 — для категории *[illegible]* и т.д.

Альтернативная гипотеза попросту состоит в том, что среднее, по крайней мере, для одной генеральной совокупности отличается от остальных. И неважно, какими именно словами записана альтернативная гипотеза. Если мы отвергаем нулевую гипотезу, то ДА не позволяет выявить, какое именно среднее (средние) значимо отличается от остальных. В последнем разделе данной главы мы кратко рассмотрим некоторые более сложные критерии, способные идентифицировать те пары средних, которые существенно различаются. Итак

(H_1: среднее, по крайней мере, для одной генеральной совокупности отличается от остальных)

Практический пример 10.1

В большом университете был [illegible] начальной биологии. Одна гру[illegible] ду "лекции плюс лаборатор[illegible] бораторные работы плюс [illegible]

... "видеозаписи лекции плюс демонстрации", причем здесь студенты могли смот- [illegible]

[illegible]

... сдавали экзамен. Были получены случайные выборки оценок для каждой груп- [illegible]

Оценки на экзамене в зависимости от метода обучения

Лекции		Демонстрации		Видеозаписи	
X	X^2	X	X^2	X	X^2
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
60	3 600	62	3 844	60	3 600
63	3 969	67	4 489	61	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
79	6 241	82	6 724	71	5 041
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
$\sum x = 636$		651		588	

[illegible]

$$SST = \sum X^2 - N\bar{X}^2$$

[illegible]

$$SSW = SST - SSB$$

[illegible]

$$\text{Внутригрупповой средний квадрат} = \frac{SSW}{dfw} = \frac{3776{,}51}{24}$$

[illegible]

$$F = \frac{120{,}41}{15[illegible]}$$

$$F = 0{,}77$$

[illegible] ному предмету не изменяются значимо в зависимости от метода обучения.

Шаг 3. Выбор распределения выборочной статистики и задание критической области

[illegible] же как и для таблицы для t-распределения, [illegible] ла степеней свободы — но уже не от одной, а от двух. Для ДА есть [illegible] степени свободы — по одной на каждую оценку дисперсий [illegible] сти соответственно. Так, цифры [illegible] степени свободы, связанные с оценкой ме[illegible] расположенные в крайнем левом [illegible] $dfw = (N-k) = 12$ (см. формулы 10.5 и 10.6) [illegible]

[illegible]

$$
\begin{aligned}
\text{[illegible]спределение выборочных статистик} &= F\text{-распределение,}\\
\text{Альфа} &= 0{,}05;\\
\text{[illegible]} &= 12\\
\text{[illegible] степеней свободы (между категориями)} &= [illegible] = 3\\
\text{[illegible]} &= 3{,}4[illegible]
\end{aligned}
$$

[illegible] внимательно посмотрите на две таблицы для F-распределения из при[illegible] больше 1,00 [illegible] что критерий, применяемый в ДА является односторонним, и нас интересуют только такие случаи, в которых дисперсия между категориями превышает дисперсию внутри категорий. Ведь F значения, меньшие 1,00, говорили бы о том, что дисперсия между категориями меньше дисперсии внутри категорий — а т.к. мы всегда заведомо не сможем отвергнуть нашу нулевую гипотезу в таких случаях, то мы просто игнорируем этот класс случаев.

Шаг 4. Вычисление статистики критерия

Это мы уже проделали в предыдущем разделе этой главы. Там мы получили наблюдаемое (экспериментальное) значение F, равное 2,27.

Шаг 5. Принятие решения

Сравним статистику критерия с критическим значением

$$
\begin{aligned}
F\,(\text{экспериментальное}) &= 2{,}27,\\
F\,(\text{критическое}) &= 3{,}49
\end{aligned}
$$

Так как статистика критерия не попала в критическую область, то наше решение таково: отвергнуть нулевую гипотезу нельзя. Итак, делаем вывод: поддержка смертной казни не различается значимо в зависимости от религии, и наблюдаемое нами различие в выборочных средних не является существенным.

10.7 Дополнительный пример на применение ДА, иллюстрирующий расчет статистики критерия и применение критерия

В этом разделе мы разберем еще один пример, иллюстрирующий [illegible] расчет статистики критерия и [illegible]. Таким образом, сначала мы еще раз вы[illegible] наблюдаемое значение F [illegible] в разделе 10.4 [illegible]

[illegible] эффективность, с которой каждый из трех [illegible] некоторую программу. В частности, нас инте[illegible] работают с документами [illegible] данной программы. Исследователь [illegible] требуемых для обработки документов, у него

есть случайная выборка из 10 наблюдений для каждого агентства. Спрашивается: существует ли значимое различие между тремя агентствами по скорости обработки документов? Имеющиеся данные приведены в таблице 10.4, в которой также [illegible] информация, которая потребуется нам для расчета статистики критерия.

Найдем SST по формуле 10.10, удобной для вычислений:

$$
\begin{aligned}
SST &= \sum X^2 - N\bar{X}^2,\\
SST &= (524 + [illegible]816 + 2462) - 30(11{,}67)^2\\
SST &= 4802 - 30([illegible]36{,}19);\\
SST &= 4802 - 4085{,}70\\
SST &= [illegible]
\end{aligned}
$$

Найдем SSB по формуле 10.[illegible]

$$
\begin{aligned}
SSB &= \sum N_k(\bar{X}_k - \bar{X})^2\\
SSB &= (10)(7 - 11{,}67)^2 + 10(13 - 11{,}67)^2 + 10(15 - 11{,}67)^2\\
SSB &= 10(-4{,}67)^2 + 10(1{,}33)^2 + 10(3{,}33)^2\\
SSB &= 10(2[illegible]81) + 10(1{,}77) + 10(11{,}09);\\
SSB &= 2[illegible]8{,}09 + 17{,}69 + 110{,}89;\\
SSB &= 346{,}7
\end{aligned}
$$

Таблица 10.4. Количество дней, затрачиваемых на обработку документов клиента для трех агентств (данные вымышленные)

	Агентство A		Агентство B		Агентство C	
Клиент	X	X^2	X	X^2	X	X^2
[illegible]	5	25	[illegible]	44	8	[illegible]
2	7	49	10	[illegible]00	8	64
3	8	6[illegible]	9	361	2	44
[illegible]	[illegible]	[illegible]00	20	400	5	[illegible]25
5	[illegible]	[illegible]6	2	144	2[illegible]	40[illegible]
6	[illegible]	[illegible]8	[illegible]	2	2	44
7	6	36	3	69	20	4[illegible]0
8	9	8[illegible]	4	[illegible]36	19	36
9	6	36	0	[illegible]00	5	225
[illegible]0	6	36	9	[illegible]	[illegible]	[illegible]
$\sum X$	[illegible]		130		150	
$\sum X^2$		524		[illegible]816		2462
$\bar{X}_k$	[illegible]		[illegible]		[illegible]	

$\bar{X}$ = [illegible]

[illegible] SSW [illegible]

$$
SSW = SST - SSB
$$

$$SSW = 716{,}30 - 346{,}7$$

$$SSW = 369{,}6.$$

[illegible] степени свободы по формулам 10.5 и 10.6 соответственно:

$$dfw = N - k = [illegible]$$

$$dfb = k - [illegible]$$

Оценки дисперсий генеральной совокупности найдем по формулам 10.7 и 10.8:

$$\text{Внутригрупповой средний квадрат} = \frac{SSW}{dfw} = \frac{369{,}66}{27} = 13{,}69$$

$$\text{Межгрупповой средний квадрат} = \frac{SSB}{dfb} = [illegible]$$

F-отношение (по формуле 10.9) составит

$$F\ (\text{экспериментальное}) = \frac{\text{Межгрупповой средний квадрат}}{\text{Внутригрупповой средний квадрат}}$$

$$F\ (\text{экспериментальное}) = \frac{[illegible]}{13{,}69}$$

$$F\ (\text{экспериментальное}) = 12{,}66$$

Теперь мы можем проверить это значение [illegible].

Шаг 1. Принятие предположения

[illegible]

[illegible]

Дисперсии генеральных совокупностей равны между собой.

Исследователь всегда сможет доказать [illegible] третье [illegible] ДА [illegible] объемы выборок приблизительно одинаковы.

Шаг 2. Формулирование нулевой гипотезы

$$H_0: \mu_1 = \mu_2 = \mu_3$$

[illegible] генеральной совокупности отличается [illegible]

Шаг 3. Выбор распределения выборочных статистик и задание критической области

[illegible] F-распределение;

альфа 0,05.

[illegible] степеней свободы [illegible] $(N - k) = [illegible] = 27$

[illegible] степеней свободы между категориями $= (k - 1) = 3 - 1 = 2$

$$F\ (\text{критическое}) = 3{,}35$$

Шаг 4. Вычисление статистики критерия

Мы уже получили наблюдаемое значение F, равное 12,66.

Шаг 5. Принятие решения

Сравним статистику критерия с критическим значением:

$$F\ (\text{критическое}) = 3{,}35;$$

$$F\ (\text{экспериментальное}) = 12{,}66$$

Так как статистика критерия попала в критическую область, то мы должны отвергнуть нулевую гипотезу об отсутствии различий. Итак, делаем вывод: вряд ли различия между тремя агентствами вызваны только случайными отклонениями. Агентства статистически значимо различаются по скорости обработки документов и [illegible] определения того, подходит ли потенциальный клиент под условия данной программы. *Чтобы попрактиковаться в проведении анализа ДА, решите задачи [illegible]*

10.8. Ограничения на применение ДА

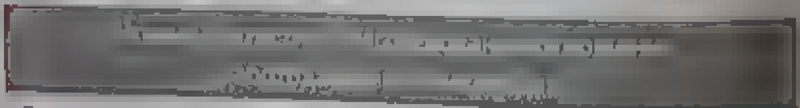
ДА подходит для случаев, когда вы хотите проверить значимость различия между [illegible] ДА называется однофакторным дисперсионным анализом [illegible]

Важное ограничение на применение ДА [illegible] приблизительно одинаковое [illegible] других переменных [illegible] может создать проблемы [illegible] ных схем формирования [illegible]. Все эти ограничения [illegible] жет мириться с некоторым [illegible]

[illegible] в том, что такие ограничения [illegible] как [illegible] при [illegible] так и при вынесении суждений об [illegible] исследований [illegible] исследований [illegible]

[illegible] ДА [illegible] для всех видов проверки [illegible] с целью выявления неслучайных различий [illegible] значимые различия [illegible] как раз и рассмотрим статистический метод, позволяющий напрямую оценить важность результатов.

Наконец, третье ограничение на применение ДА появляется из-за [illegible] гипотезы. Помните, что в результате отклонения нулевой гипотезы принимается [illegible] ограничение состоит в том, что альтернативная гипотеза неконкретна: она просто говорит о том, что среднее, по крайней мере, для одной генеральной совокупности отличается от остальных. Очевидно, что нам хотелось бы узнать, какое (какие) именно из различий значимо (значимы). На этот вопрос [illegible] в таблице 10.4 достаточно очевидно, что агентство А является источником [illegible] большего различия. Однако такой неформальный метод (на глазок) может привести к [illegible] выводам, поэтому мы должны быть осторожны, делая выводы о том, какие средние значимо отличаются от остальных. В следующем разделе мы рассмотрим подробный пример такой интерпретации ДА и опишем апостериорные критерии, называемые также анализом post hoc или after the fact (после того, как [illegible]) [illegible] для апостериорных критериев выходят за рамки нашего [illegible]



Ситуация с занятостью инвалидов достаточно мрачная: только 29% из них имеют работу, по сравнению с 79% занятости для [illegible] [illegible] Для взрослых [illegible] [illegible] имел физическую инвалидность (травму мозга, прикововавшую его к инвалидному креслу), третий кандидат [illegible]

[illegible]

[illegible] 246 [illegible] [illegible] профессионализму. Таблица 1 показывает, что ожидания исследователей подтвердились.

Таблица 1. Результаты однофакторного дисперсионного анализа рейтингов пригодности к найму

	Не инвалид	Физическая инвалидность	Психическая инвалидность	ДА F-отношение	ДА Альфа
[illegible] (идеальных рейтинг)	74,00	83,00	86,00	17,97	p<0,000
[illegible]	139,59	125,95	120,15		
[illegible]	17,04	17,54	15,04		

Среднее для кандидата, не имеющего инвалидности, было наибольшим, а средние для двух остальных кандидатов, имеющих инвалидности, были близки между собой. Исследователи применили апостериорный критерий и подтвердили, что [illegible] существует значимое различие между кандидатами-неинвалидами и кандидатами с физической инвалидностью, а также между кандидатами-неинвалидами и кандидатами с психической инвалидностью. Рейтинги пригодности к найму для кандидатов с физической инвалидностью и кандидатов с психической инвалидностью не различаются значимо". Авторы делают вывод, что шансы получить работу для инвалидов — "большинства [illegible] в наиболее невыгодном финансовом положении" — меньше, чем у неинвалидов. Обсуждаются некоторые стратегии по улучшению этой ситуации для социальных работников, инвалидов и других обеспокоенных этой проблемой.

[illegible] C. Bricout and Kia J. Bentley. 2000. "Disability Status and Perception of Employability by [illegible]" [illegible] Work Research [illegible]

10.9. Интерпретация статистик: меняется ли сексуальная активность в зависимости от семейного положения?

Меняется ли сексуальная активность респондента в зависимости от его (ее) семейного положения? Занимаются ли состоящие в браке сексом чаще, чем не состоящие [illegible] [illegible] (General Social Survey) [illegible] выборку (см. главу 6) из населения США.

[illegible] переменные: семейное положение, количество сексуальных партнеров за прошлый год [illegible] относится к номинальной шкале [illegible]

[illegible]

имеет 5 категорий, это будет [illegible] переменная. [illegible]

[illegible] **каждого интервала на его среднюю точку** [illegible]

[illegible] указали очень боль[illegible]

Таблица 10.5. Описательные статистики для переменных, характеризующих сексуальную активность

	Количество сексуальных партнеров за последние 12 месяцев	Количество раз занятий сексом за последние 12 месяцев
Среднее	1,23	55,30
Стандартное отклонение	3,41	67,62
N	2403	2320

Средние и стандартные отклонения для переменной *количество сексуальных партнеров* [illegible] это является значимым на уровне ме[illegible]

[illegible]

количество раз занятий сексом за последние 12 месяцев в зависимости от семейного положения [illegible]

[illegible] (63,8 раз в год). Никогда не состоявшие в браке [illegible], что занимаются сексом около 59 раз в год. [illegible] сообщили более низкую цифру [illegible] влиянию семейного положения.

Таблица 10.6. Количество сексуальных партнеров, в зависимости от семейного положения

	Семейное положение					ДА	
	В браке	Вдовец (вдова)	Официально разведен(а)	Официально в браке, но фактически в разводе	Никогда не состоял(а) в браке	F-отношение	Альфа
Среднее	1,08	0,26	1,31	1,37	1,83	9,84	<0,05
Стандартное отклонение	2,94	0,53	3,38	1,41	4,76		
N	1,157	224	365	82	675		

Таблица 10.7. Количество раз занятий сексом за последние 12 месяцев, в зависимости от семейного положения

	Семейное положение					ДА	
	В браке	Вдовец (вдова)	Официально разведен(а)	Официально в браке, но фактически в разводе	Никогда не состоял(а) в браке	F-отношение	Альфа
Среднее	63,83	12,47	48,23	58,79	58,58	27,960	<0,05
Стандартное отклонение	62,66	39,72	69,29	78,80	75,69		
N	102	209	364	81	564		

Таблицы 10.6 и 10.7 [illegible], какие именно различия из этих таблиц наиболее важны и вносят наибольший вклад в значимость F-отношения. Мы можем определить на глаз, что [illegible] отличается от всех остальных групп. Может быть, [illegible] тоже важны? Апостериорные критерии [illegible] выборочных средних дают объективный ответ [illegible]

Мы не будем описывать вычислительные [illegible] интерпретации результатов этих критериев [illegible]. Кратко, то апостериорные критерии [illegible] категорий (например, состоящ[illegible] [illegible] более одной пары [illegible] первого рода (альфа, т.е. вероятность [illegible]

[illegible] критерий [illegible] значимых различий [illegible] критерии [illegible] [illegible] значимость [illegible] [illegible] Элементами [illegible] среднего [illegible] количество партнеров [illegible] [illegible] 0,8161 [illegible] значение показано на пересечении двух названных категорий. Звездочка (*) рядом со значением означает, что данное различие является значимым на уровне 0,05.

Таблица 10.8. Апостериорный тест показывает различия в сексуальной активности в зависимости от семейного положения

	Различия в средних для количества партнеров					
Семейное положение	Средние группы категории	В браке	Вдовец (вдова)	Официально разведен(а)	Официально в браке, но фактически в разводе	Никогда не состоял(а) в браке
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]			[illegible]	[illegible]	[illegible]
[illegible]	[illegible]				[illegible]	[illegible]
Официально в браке, но фактически в разводе	[illegible]					[illegible]
Никогда не состоял(а) в браке	[illegible]					
	Различия в средних для количества раз занятий сексом					
Семейное положение	Средние группы категории	В браке	Вдовец (вдова)	Официально разведен(а)	Официально в браке, но фактически в разводе	Никогда не состоял(а) в браке
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]			[illegible]	[illegible]	[illegible]
[illegible]	[illegible]				[illegible]	[illegible]
Официально в браке, но фактически в разводе	[illegible]					[illegible]
Никогда не состоял(а) в браке	58,56					

* [illegible] является значимым на уровне 0,05; модифицированный критерий [illegible]

[illegible] — Прим. науч. ред.

В верхней части таблицы 10.8 приведены различия в средних для количества сексуальных партнеров. [illegible] оказывается, что наименьшее [illegible] от всех остальных групп. Кроме этого, [illegible] группа [illegible] в браке имеет значимо больше партнеров, чем группа состоящих в браке и группа разведенных соответственно. В нижней части таблицы [illegible] различия в средних для количества раз занятий сексом. И здесь группа [illegible] значимо отличается от всех остальных групп. Кроме этого, критерий нашел еще два менее очевидных различия: [illegible] занимаются сексом значимо реже, чем состоящие в браке [illegible] состоящие в браке [illegible] вдовевших наиболее отличается от других [illegible] 10.6 и 10.7, но указанные выше дополнительные значимые различия могли ускользнуть от нашего внимания, если бы мы не применили апостериорный критерий.

Итак, сформулируем выводы. Дисперсионный анализ показал, что семейное положение приводит к значимому различию в уровнях сексуальной активности взрослых американцев. Индивидуумы, состоящие в браке, занимаются сексом значимо чаще со значимо меньшим количеством партнеров. Люди, никогда не состоявшие в браке, имеют значимо большее количество сексуальных партнеров, чем состоящие в браке или официально разведенные индивидуумы. Овдовевшие люди характеризуются самыми низкими показателями по обеим мерам сексуальной активности.

Заключение

1. Однофакторный дисперсионный анализ — это мощный метод для проверки значимости различий, применяемый повсеместно, когда нас интересует сравнение более чем двух категорий или выборок. Вероятно, проще всего представить себе анализ ДА как дальнейшее развитие t-критерия, проверяющего значимость различия между средними двух выборок.
2. В дисперсионном анализе уровень изменчивости между категориями сравнивается с уровнем изменчивости внутри категорий. Если нулевая гипотеза [illegible] различий ложна, то тогда должна наблюдаться сравнительно большая изменчивость между категориями и сравнительно малая изменчивость внутри категорий. Чем больше [illegible] между категориями [illegible] ям внутри этих категорий, тем больше вероятность того, что мы сможем отвергнуть нашу нулевую гипотезу.
3. [illegible] ДА [illegible] широко распространенные и общедоступные [illegible] SPSS, — с помощью которых [illegible]

ваться такими программами, то пусть стремление облегчить себе проведение ДА станет вашим стимулом.

[illegible]

критическим значением F-отношения (определяемого для заданного уровня значимости альфа и вычисленных двух чисел степеней свободы). Если принято ре[illegible]

[illegible]

Сводка формул

Общая сумма квадратов

$$SST = \sum [illegible]$$

Две компоненты общей суммы квадратов

$$SST = [illegible]$$

[illegible]

0.3 $$SSW = \sum (X_i - \bar{X}_k)^2$$

Межгрупповая сумма квадратов

[illegible]

Число степеней свободы для SSW

10.5 $$dfw = N - k$$

Число степеней свободы для SSB

[illegible]

[illegible]

0.7 $$\frac{SSW}{[illegible]}$$

Внутригрупповой средний квадрат

10.8 $$\frac{SSB}{dfb}$$

F-отношение

10.9 $$F = \frac{\text{Межгрупповой средний квадрат}}{\text{Внутригрупповой средний квадрат}}$$

[illegible] SST

$$SST = \sum [illegible]$$

[illegible] SSW [illegible]

$$SSW = SST - SSB$$

Словарь

Апостериорный критерий (post hoc test). [illegible]

Внутригрупповая сумма квадратов (SSW). [illegible]

ДА. [illegible]

Дисперсионный анализ. Метод, используемый для проверки значимости различий [illegible] двух выборок.

Межгрупповая сумма квадратов (SSB). [illegible] общего среднего.

Общая сумма квадратов (SST). Сумма квадратов отклонений значений (наблюдений) [illegible]

Однофакторный дисперсионный анализ. [illegible] влияние одной независимой переменной [illegible] (отклик).

Оценка среднего квадрата. [illegible] (SSB) [illegible]

F-отношение. [illegible]

Мультимедийные ресурсы

[illegible]

[illegible]

[illegible] широкий спектр ресурсов для [illegible] разделе *Student Resources* этого сайта. [illegible]

- Обзор основ математики,
- [illegible]
- Обучающие карточки
- Ссылки на полезные интернет-ресурсы,
- Дополнительные задачи по главам,
- Таблица случайных чисел;
- Примеры и упражнения с использованием пакетов статистической обработки данных MicroCase и SPSS
- [illegible]
- Проверка гипотез для переменных, измеряемых по порядковой шкале

Задачи

10.1 Проведите ДА для каждого из 3-х наборов баллов, приведенных ниже. *Подсказка: следуйте вычислительной схеме, приведенной в разделе 10.4; записывайте все суммы и средние в таблицы, наподобие таблиц 10.3 и 10.4.*

а.

Категория		
A	B	C
[illegible]	[illegible]	[illegible]
7	12	6
8	14	18
[illegible]	[illegible]	[illegible]

б.

Категория		
A	B	C
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	7	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]		

в.

Категория			
A	B	C	D
[illegible]	[illegible]	5	5
[illegible]	[illegible]	8	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	5	[illegible]

10.2. SOC

Люди какого типа больше других вовлечены в организации, объединения [illegible] или общину? Человек какого [illegible] [illegible] PTA [illegible] [illegible] являетесь." и некоторые другие. Какие из различий являются значимыми?

а. Членство в зависимости от образования

Образование меньшее чем школа	Школа	Колледж
0	1	0
1	3	3
2	3	4
3	4	4
4	5	4

б. Членство в зависимости от длительности проживания в своей общине

Менее 2-х лет	2–5 лет	Более 5-ти лет
0	0	1
1	2	3
3	3	3
4	4	4
4	5	4

в. Членство в зависимости от степени увлечения просмотром телевизора

Малая или отсутствует	Средняя	Высокая
[illegible]	[illegible]	4
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	4
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	5

г. [illegible]

Детей нет	Один ребенок	Более чем один ребенок
[illegible]	2	0
[illegible]	3	3
[illegible]	4	[illegible]
[illegible]	4	4
[illegible]	4	[illegible]

10.3. SOC

Среди населения определенной местности [illegible] объемом 18 семейных пар [illegible] деления власти и ответственности [illegible]

[illegible] ...ированы по типу отношени[illegible] [illegible] ...ский брак (пар[illegible]) [illegible] работает ли жена). Изменяется ли значимо ответ[illegible] — независимо от того, работает ли жена) [illegible] ответственность за принятие решений или счастье в браке в зависимости от типа [illegible]

а. Ответственность за принятие решений

Традиционная пара	Двухкарьерная пара	Гражданский брак
7	8	2
8	5	1
[illegible]	[illegible]	3
[illegible]	[illegible]	[illegible]

б. [illegible]

Традиционная пара	Двухкарьерная пара	Гражданский брак
[illegible]	[illegible]	[illegible]

10.4. CJ

В г. [illegible] реализовывались две отдельные программы по снижению преступности. Причем в некоторых районах реализовывалась первая программа, в других — вторая, а в остальных никакая программа не реализовывалась. Первая — это программа наблюдения за соседями, в которой граждане активно вовлекаются в процесс предотвращения преступлений. Вторая программа заключается в патрулировании полицией своего района — причем не на патрульных машинах, а пешком. Были ли эти две программы успешными (в показателях уменьшения количества преступлений, в процентах, по данным полиции) [illegible] выборкам для всего города.

Прирост количества преступлений (в процентах) по сравнению с прошлым годом

Первая программа (наблюдение за соседями)	Вторая программа (пеший патруль)	В отсутствие программ
-2	[illegible]	+3
[illegible]	15	[illegible]
[illegible]	6	[illegible]
[illegible]	10	[illegible]
+7	[illegible]	+8
+0	[illegible]	[illegible]

10.5. SEX

Правда ли, что сексуально активные подростки лучше информированы о СПИДе и других потенциальных проблемах, связанных с сексом, чем неактивные? [illegible] трех случайных выборок подростков: сексуально неактивных, активных и имеющих только одного партнера (постоянного партнера), активных и имеющим более одного партнера — проведен опрос из 15 пунктов на тему общих [illegible] о сексе и здоровье. Существуют ли значимые различия в результатах опроса (баллах)?

Сексуально неактивные подростки	Активные и имеющие только одного партнера	Активные и имеющие более одного партнера
0	11	12
2	11	12
8	6	10
10	5	4
8	15	9
5	10	15

10.6. PS

Изменяется ли значимо процент явки избирателей в зависимости от типа выборов? Случайная выборка из избирательных участков показывает следующую явку избирателей в зависимости от типа выборов. Оцените приведенные данные на значимость различий.

Только местные выборы	Данного штата	Национальные
[illegible]	35	42
[illegible]	56	40
[illegible]	35	52
[illegible]	40	66
[illegible]	45	78
[illegible]	42	62
[illegible]	65	57
[illegible]	62	75
[illegible]	25	72
[illegible]	47	51
[illegible]	[illegible]	[illegible]
[illegible]	56	59

10.7. GER

Теряют ли пожилые граждане интерес к [illegible] случайных выборок респондентов — [illegible] ответственно — проведен [illegible]

Ниже [illegible]

Школьники 15–18	Молодежь 21–30	Люди среднего возраста 40–55	Пенсионеры 65+
0	0	[illegible]	6
1	0	3	6
1	2	3	6
2	2	4	7
2	4	4	7
2	4	5	8
3	4	6	10
5	6	7	10
5	7	7	10
7	7	8	10
7	7	8	10
9	10	10	10

10.8 SOC

Из базы данных Общенационального социологического обследования (General Social Survey) была отобрана малая случайная выборка респондентов. Респонденты были разделены по следующим категориям: городской житель, житель пригорода, сельский житель. Есть ли статистически значимое различие в зависимости от места проживания по каждой переменной, перечисленной ниже?

а. Престиж профессии (PRESTG80)

Городской житель	Житель пригорода	Сельский житель
32	40	30
45	48	40
42	50	40
47	55	45
48	55	[illegible]
50	60	[illegible]
51	65	52
55	70	55
6[illegible]	75	[illegible]
6[illegible]	75	6[illegible]

б. Количество детей (CHILDS)

Городской житель	Житель пригорода	Сельский житель
[illegible]	0	1
[illegible]	1	4
5	6	2
2	0	3
[illegible]	2	3
[illegible]	[illegible]	3
2	[illegible]	2
[illegible]	[illegible]	5
[illegible]	[illegible]	6

Городской житель	Житель пригорода	Сельский житель
1	2	4
3		6

в. Доход семьи (INCOME86)

Городской житель	Житель пригорода	Сельский житель
5	6	5
7	8	5
8	1	[illegible]
1	12	9
8	2	7
4	1	[illegible]
8	[illegible]	9
	9	7
7	9	4
10	2	6

г. Посещение церкви (ATTEND)

Городской житель	Житель пригорода	Сельский житель
0	6	[illegible]
[illegible]	0	5
0	2	4
4	5	4
5	8	9
8	5	4
[illegible]	8	8
5	7	8
7	2	8
4	6	5

д. Количество часов в день, проводимых за просмотром телевизора (TVHOURS)

Городской житель	Житель пригорода	Сельский житель
5	5	3
[illegible]	7	7
2	0	5
3	2	0
0	3	[illegible]
1	6	3
4	1	5
4	3	10
5	4	3
9		

Использование *SPSS for Windows* для проведения дисперсионного анализа

Демонстрационный пример с использованием SPSS 10.1

Увеличивается ли с возрастом политический консерватизм?

[illegible] на самом деле принадлежат к порядковой шкале (например доход), дру- [illegible] [illegible] (например возраст). Поэтому, чтобы было что демонстрировать, я допус- [illegible]

Давайте начнем с исследования идеи о том, что в обществе США политическая идеология связана с возрастом. Часто говорят, что люди с возрастом становятся более консервативными в отношении широкого ряда вопросов, это может привести [illegible]

[illegible] переменной указывают на больший уровень консерватизма. Чтобы можно было применить ДА, нам придется преобразовать переменную [illegible] [illegible] делиться на категории, я запустил команду Frequencies (Частоты) и нашел так[illegible] [illegible] что три группы примерно одинак[illegible]

[illegible] Analyze [illegible] Compare Means [illegible] Anova [illegible]

[illegible] стрелке, чтобы переместить имя этой переменной в список зависимых переменных Dependent List. Заметим, что вы можете выбрать более [illegible]

[illegible] (например, с име- [illegible] стрелке, чтобы переместить эту переменную в [illegible] [illegible] независимой переменной)

[illegible] Options (Опции) [illegible]

[illegible] они будут вычислены в [illegible]

[illegible] результатов [illegible]

Описательные статистики, самооценка респондента как либерала или консерватора

			Стандартное отклонение	Ошибка	95% доверительный интервал для среднего: нижний	верхний	[illegible]	[illegible]
			1.36	6.62E-02	3.78	4.04	1	7
			1.40	6.38E-02	3.91	4.16	1	7
			[illegible]	6.48E-02	[illegible]	4.40	1	7
[illegible]	3.17	4.08	[illegible]	3.6[illegible]E-02	[illegible]	4.15	1	7

[illegible] F [illegible]

[illegible]

Демонстрационный пример с использованием SPSS 10.2. Изменяется ли политическая идеология в зависимости от социального класса?

[illegible]

В [illegible] GSS есть несколько переменных, каждую из которых можно [illegible] *class*, *prestg80*. В реальном социологическом исследовании нам бы, наверное, при[illegible] ограничимся только переменной *class* (метка которой гласит SUBJECTIVE CLASS IDENTIFICATION — Личная классовая идентификация). Следуйте указаниям по проведению **One-way Anova**, описанным в демонстрационном примере 10.1. Как и ранее, укажите переменную *polviews* как зависимую, но в роли фактора (независимой переменной) укажите переменную *class*. Окно результатов будет иметь следующий вид.

Описательные статистики: самооценка респондента как либерала или консерватора

	N	Среднее	Станд. отклонение	Ошибка	95% доверительный интервал на среднее: нижний	верхний	Минимум	Максимум
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Высшие [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

ANOVA, самооценка респондента как либерала или консерватора

	Сумма квадратов	Ст. св.	Средний квадрат	F	Знач.
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	1306	[illegible]		
[illegible]	[illegible]	[illegible]			

[illegible] результат в ANOVA [illegible] вероятность [illegible] результатов [illegible] 0,05. Посмотрите на средние групп [illegible] классы являются наиболее либеральными [illegible] балла класса. Полученный результат [illegible] 0,05 политический консерватизм [illegible]

Демонстрационный пример с использованием SPSS 10.3. Второй пример на выявление различий в отношении к абортам

В демонстрационном примере 9.2 мы обнаружили, что переменная *sex* оказывает статистически значимое влияние на уровень абортов, переменную *abscale* (причем отличие статистики от критического значения небольшое). А теперь давайте исследуем влияние вероисповедания респондента (переменная *relig*) на переменную *abscale*. Щелкните **Analyze** (Анализ), **Compare means** (Сравнить средние), а затем — **One-way Anova** (Однофакторный дисперсионный анализ). Появится окно **One-way Anova**. Укажите *abscale* в качестве зависимой переменной, а *relig* — в качестве фактора. При необходимости см. демонстрационный пример 9.2, где рассказывается, как вычислить переменную *abscale*. Окно результатов будет иметь следующий вид.

Описательные статистики: ABSCALE

	N	Среднее	Станд. отклонение	Ошибка	95% доверительный интервал на среднее: нижний	верхний	Минимум	Максимум
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	227	2,7665	0,7059	4,685E-02	2,6742	2,8588	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

ANOVA: ABSCALE

	Сумма квадратов	Ст. св.	Средний квадрат	F	Знач.
[illegible]	[illegible]	4	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	0,428		
Общая	384,090	846			

[illegible]

Апостериорные критерии (post hoc analysis)[3]

Применив апостериорные критерии, мы сможем [illegible] какими парами средних являются значимыми. [illegible] SPSS в окне One-way Anova [illegible] выполните команду **One-way Anova** для переменной *relig* [illegible] Post Hoc [illegible] в качестве зависимой переменной), щелкнув [illegible] критерии) в этом же окне One-way Anova [illegible]

[illegible]

[illegible] Эта таблица довольно длинная, поэтому здесь мы ее не [illegible] [illegible]о знать, что нужная информация находится в столбце с [illegible] Разность между средними). В этом столбце приведены [illegible] значениями двух категорий независимой переменной (по всем [illegible]рий). Например, в верхней строке сравниваются протестан[illegible] [illegible]ется, что разность их средних равна 0.003 (это значение бу[illegible] в формате с плавающей запятой ("в научном формате" SPSS), как [illegible] его перевода в привычный вид переместите десятичную точку на три [illegible] В следующей строке сравниваются *протестанты* и *иудеи* и сообща[illegible] [illegible] Различия, являющиеся статистически значимыми при уровне значимости 0.05, помечаются звездочкой (*). Из этой таблицы [illegible] мы делаем вывод, что большинство статистически значимых различий наблюдается между *протестантами* и *католиками* с одной стороны, и *иудеями* и [illegible] с другой стороны.

[illegible]пражнения

10.1 Как продолжение демонстрационного примера SPSS и MicroCase 10.2 проведите ДА с независимыми переменными (факторами) *income91*, *prest80*, *papres* для зависимой переменной *polviews*. Демонстрируют ли эти меры принадлежности к социальному классу тот же тип связи с *polviews*, что и переменная *class*? Чтобы выполнить это упражнение, сначала перекодируйте каждую из названных независимых переменных, чтобы получить три категории. Для каждой переменной запустите Frequencies (для SPSS) или Univariate (для Microcase) и найдите точки, разделяющие выборку на три группы примерно одинакового объема.

10.2 Какая еще переменная (кроме *sex* и *religion*) может иметь статистически [illegible] [illegible] кодируйте ваши независимые перем[illegible] четыре категории.

11 Проверка статистических гипотез IV: Критерий хи-квадрат

Цели изучения

1. Идентифицировать и приводить примеры ситуаций, в которых [illegible] применение критерия хи-квадрат.
2. Объяснить структуру двухвходовой таблицы сопряженности [illegible] и концепцию независимости в применении к ожидаемым и наблюдаемым частотам такой таблицы.
3. Объяснить логику процесса проверки статистических гипотез применительно к двухвходовой таблице сопряженности.
4. Применять критерий хи-квадрат, следуя пятишаговому алгоритму процесса принятия решений. Правильно интерпретировать полученные результаты.
5. Объяснить ограничения применения критерия хи-квадрат, а особенно — в чем отличие между статистической значимос[illegible]

11.1. Введение

[illegible]

[illegible] как [illegible] раздела 9.4 [illegible] критерий [illegible] нормальности [illegible]

[illegible] Решение об отклонении [illegible] модели, или же нулевая гипотеза, — неверны. Конечно, обычно при этом мы считаем [illegible] нулевая гипотеза неверна и отвергаем ее. Чем более уверены [illegible] уверенность в том, что неверны [illegible] Слабость [illegible] новых допущений модели означает, что наше решение об отклонении нулевой [illegible] уверенностью.

[illegible] критерия независимости хи-квадрат [illegible] переменными [illegible] принадлежащими к любым [illegible] ном, а вот критерий независимости хи-квадрат легко справится с такой более сложной переменной.

11.2 Двухвходовые таблицы сопряженности

Критерий независимости хи-квадрат применяется к двухвходовым таблицам сопряженности [illegible] потому что в них представляются результаты наблюдений двух различных переменных одновременно. Двухвходовые таблицы сопряженности [illegible] [illegible] обеих переменных.

[illegible]

делится на категории белые и чернокожие, а уровень вовлеченности — [illegible] высокий и низкий.

Общепринято, чтобы категории независимой переменной (т.е. переменной, используемой для предсказания значений [illegible] объясняется зависимая переменная) [illegible] в столбцах таблицы сопряженности, а категории зависимой переменной — в строках таблицы сопряженности. В нашем примере независимой переменной является раса, поэтому в столбцах мы представили ее категории, а в строках — категории переменной *уровень вовлеченности* (высокий и низкий соответственно). Таблица 11.1 — это пустая двухвходовая таблица сопряженности для выборки объемом в 100 наблюдений, которую нам предстоит заполнить. Отметим, что эта таблица имеет шапку, в которой указаны имена переменных: сначала зависимой, потом независимой. Понятно, что аналогичные шапки должны сопровождать все таблицы, графики и схемы.

Таблица 11.1. Уровень вовлеченности в добровольные организации в зависимости от расовой группы для 100 пожилых граждан

	Расовая группа		
Уровень вовлеченности	*Чернокожие*	*Белые*	
Высокий			50
Низкий			50
	50	50	100

Чтобы заполнить таблицу 11.1, мы должны классифицировать каждого респондента [illegible] в терминах расы и степени вовлеченности [illegible] (встречаемость) каждой комбинации расы и степени вовлеченности [illegible] возможных комбинаций четыре [illegible] чернокожих граждан, имеющих высокую степень вовлеченности [illegible] степени вовлеченности — в нижнюю правую ячейку и т.д. После того как мы [illegible] делить наблюдения по ячейкам, в каждой ячейке будет показана наблюдаемая встречаемость соответствующей комбинации. В таблицах сопряженности [illegible] подсчитываются еще и частные суммы по строке и столбцу [illegible] ми частотами по строке и столбцу соответственно. Общее [illegible] таблице (N) записывается на пересечении строки [illegible] ные суммы.

В заключение отметим, что нашу таблицу сопряженности можно расширить [illegible] бы представить более сложные переменные. Так, если мы [illegible] дование больше расовых групп (например [illegible] испанского происхождения), то нам придется просто добавить [illegible]. Аналогично для зависимой переменной [illegible] пости не двумя, а тремя категориями [illegible] лось бы просто добавить к таблице [illegible] строку.

11.3. Логика применения критерия хи-квадрат

[illegible] хи-квадрат имеет несколько [illegible] [illegible] критерий согласия хи-квадрат [illegible] квадрат. Критерий согласия хи-квадрат [illegible] мы будем иметь дело с критерием независимости хи-квадрат. Мы уже [illegible] торых требуется для применения [illegible] критерия для двух выборок [illegible]. [illegible] что для применения этих методов требуются независимые [illegible] что выбор любого наблюдения одной выборки не оказывает никакого [illegible] на вероятность того [illegible] какое-то конкретное значение [illegible] выборки (см. раздел 9.2).

Применительно же к [illegible] мости имеет немного иное значение. Две переменные [illegible] [illegible] категории одной переменной [illegible] влияния на вероятность того, что это наблюдение попадет в [illegible] другой переменной. [illegible] раса [illegible] будут независимы друг от друга, если [illegible] не [illegible] этого [illegible] участия (высокая или низкая). Другими словами [illegible] симы, если степень участия и принадлежность к расе [illegible]

Таблица 11.2. Частоты наблюдений в ячейках [illegible] случае, когда участия и расовая принадлежность независимы друг от друга

Степень участия	Чернокожие	Белые
Высокая	25	25
Низкая	25	[illegible]

[illegible]

[illegible]

нулевая гипотеза ложна, то тогда эти раз[illegible] не различия между ожидаемыми (f_e) и наблюдаемыми [illegible] роятность того, что наши переменные [illegible] независимыми [illegible] [illegible] вероятность того, что мы сможем [illegible] нашу нулевую ги[illegible]

11.4. Расчет статистики для критерия независимости хи-квадрат

[illegible] критерия — χ^2 (экспериментальное) [illegible] ния статистики в распределении [illegible] χ^2 [illegible] χ^2 критическ[illegible] ределяемым по таблице распределения [illegible] уровня значимости альфа и рассчитанного числа [illegible] [illegible] общую модель проверки [illegible] тистики критерия

Формула 11.1 χ^2 [illegible]

[illegible]

Формула 11.2

[illegible]

То есть ожидаемая частота наблюдений в данной ячейке равняется общему количеству наблюдений в строке (маргинальной частоте по строке), умноженному на общее количество наблюдений в столбце (маргинальную частоту по столбцу) и [illegible] [illegible] наблюдений в таблице N.

[illegible] ожидаемые частоты, следует применить формулу [illegible] вычесть ожидаемую частоту из наблюдаемой частоты, возвести эту разность в квадрат [illegible]

Чтобы вам стало понятнее, давайте рассмотрим еще один пример (табл. 11.3). [illegible] студент по программе, аккредитованной Советом по образованию социальных работников (Council on Social Work Education) [illegible] ряды которой соответствуют столбцам таблицы; 2) принят ли данный студент [illegible] работника [illegible] соответствуют строкам таблицы.

Таблица 11.3. Трудоустройство на работу по специальности для 100 выпускников колледжа в зависимости от аккредитационного статуса программы обучения

Статус занятости	Аккредитационный статус		
	Аккредитован	Не аккредитован	Итого
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	25	35	[illegible]
	55	45	[illegible]

[illegible]

Таблица 11.4. Ожидаемые частоты для таблицы 11.3

Статус занятости	Аккредитационный статус		
	Аккредитован	Не аккредитован	Итого
[illegible]	22	18	40
[illegible]	33	27	60
	55	45	100

Заметим, что маргинальные частоты по строкам и столбцам, а также общее количество [illegible] в табл. 11.4 [illegible] равняются суммам по

строкам и столбцам для наблюдаемых частот. [illegible] правильности [illegible] ожидаемых частот [illegible]

Найдем значение статистики критерия по формуле 11.1.

$$\chi^2(\text{экспериментальное}) = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2(\text{экспериментальное}) = \frac{(30-22)^2}{22} + \frac{(10-18)^2}{18} + \frac{(25-33)^2}{33} + \frac{(35-27)^2}{27}$$

$$\chi^2(\text{экспериментальное}) = \frac{64}{22} + \frac{64}{18} + \frac{64}{33} + \frac{64}{27}$$

$$\chi^2(\text{экспериментальное}) = 2{,}91 + 3{,}56 + 1{,}94 + 2{,}37$$

$$\chi^2(\text{экспериментальное}) = 10{,}78$$

Это экспериментальное (рассчитанное по выборке) значение статистики хи-квадрат следует проверить на значимость. *Чтобы попрактиковаться в вычислении статистики для критерия независимости хи-квадрат, решите задачу 11.1.*

11.5 Критерий независимости хи-квадрат

[illegible] мы будем следовать пятишаговому алгоритму [illegible] представления процесса принятия решений [illegible] продолжаем рассматривать [illegible] представленного в таблице 11.3.

Шаг 1. Принятие предположений

Модель: независимые случайные выборки
Зависимая переменная относится к [illegible]

Шаг 2. Формулирование нулевой гипотезы

[illegible] нулевая гипотеза для критерия независимости хи-квадрат [illegible] переменные независимы [illegible] различия между наблюдаемыми и ожидаемыми частотами [illegible] Как обычно, альтернативная гипотеза [illegible] нулевой гипотезе. То есть, если мы отвергаем H_0, [illegible] нулевую гипотезу.

H_0: [illegible] переменные независимы

H_1: [illegible] переменные зависимы

Шаг 3. Выбор распределения выборочной статистики и задание критической области

Распределение [illegible] выборочных статистик хи-квадрат [illegible] от Z-распределения и t-распределения [illegible]

хвост справа от него. Поэтому для критерия мы [illegible] находится в области правого хвоста распределения выборочных статистик.

Критические значения χ^2 (критическое) [illegible] приведены в приложении В. [illegible] из таблицы для распределения уровня значимости альфа [illegible] число степеней свободы [illegible] Однако главное отличие от таблиц [illegible]

[illegible]

где df — число степеней свободы

[illegible] количество столбцов минус 1 (от "[illegible]" — [illegible]

[illegible] число степеней свободы независимо от количества наблюдений в выборке. [illegible]

[illegible] уровень значимости альфа = 0,05 [illegible] критическое значение хи-квадрат составит 3,84. Перенесем наши выводы в [illegible]

Распределение выборочных статистик = распределение хи-квадрат

[illegible] = 0,[illegible]

χ^2 (критическое) = 3,84

Распределение выборочных ст[illegible]

Рис. 11.1

Распределение хи-квадрат [illegible]

[illegible]

Шаг 4. Вычисление статистики критерия

Вычисление статистики критерия мы подробно рассмотрели [illegible]

$$\chi^2(\text{экспериментальное}) = \sum\frac{(f_o - f_e)^2}{f_e}.$$

$$\chi^2(\text{экспериментальное}) = 10{,}78.$$

Шаг 5. Принятие решения

Сравним статистику критерия с критическим значением

χ^2 (экспериментальное) = 10,7[illegible]

χ^2 (критическое) = 3,84

Мы видим, что статистика критерия попадает в критическую область, следовательно, мы отвергаем нулевую гипотезу о независимости переменных. Маловероятно, что наблюдаемые значения частот в ячейках таблицы 11.3 вызваны только случайностью. Исследуемые переменные зависимы. На основе данной выборки можно сделать вывод, что вероятность трудоустройства на работу в должности социального работника зависит от аккредитации [illegible] программы обучения. [illegible] *практиковаться в применении критерия независимости хи-квадрат и интерпретации результатов, решите задачи 11.2–11.15*

Отметим, что критерий независимости хи-квадрат, говоря о том, что наши переменные зависимы, ничего не говорит о том, какова именно связь между переменными. Больше или меньше вероятность того, что выпускники аккредитованных программ устроятся на работу в [illegible] ("аккредитованные" и "неаккредитован[illegible]) [illegible] вычисления. Из [illegible] аккредитованных программ ра[illegible] [illegible] 55% мы получили 54,55% [illegible] 55 [illegible] статистически значимой (по критерию хи-квадрат) [illegible] проценты. Делаем вывод: вероятность того, [illegible] программ устроятся на работу в должности [illegible] с выпускниками неаккредитованных программ. [illegible] таблицах сопряженности мы рассмотрим [illegible] связи для двух переменных.

11.6 Пример применения критерия независимости хи-квадрат

[illegible] мы рассматривали [illegible] более сложный пример [illegible]

[illegible] решения для таблицы, большей чем 2×2. Как вы увидите, потребуется больше [illegible] вычисления (т.к. в таблице стало больше ячеек), но в остальном все точно так же [illegible]

[illegible] состоящие в браке [illegible] семейных обязанностей [illegible] зависимость от семейного [illegible] 453 студентов [illegible] (GPA, grade point average) [illegible] "средний", "плохой". Результаты представлены в таблице 11.5.

[illegible] состоящих в браке [illegible] ожидаемая частота составит (160 × 175)/453 = 61,8. [illegible] 278 [illegible] 453 [illegible] 98,2. [illegible] маргинальные частоты, которые необходимы). Результат показан в таблице 11.6.

Теперь давайте рассчитаем χ^2 (экспериментальное). Будьте внимательны, подставляйте именно те f_e и f_o, которые нужны для данной ячейки!

Таблица 11.5. Средний балл (GPA) в зависимости от семейного положения для 453 студентов колледжа

	Семейное положение		
Средний балл (GPA)	Состоящие в браке	Не состоящие в браке	Итого
Хороший	[illegible]	[illegible]	160
Средний	60	1[illegible]0	170
Плохой	45	78	123
Итого	175	278	453

Таблица 11.6. Ожидаемые частоты для таблицы 11.5

	Семейное положение		
Средний балл (GPA)	Состоящие в браке	Не состоящие в браке	Итого
Хороший	61,8	98,2	160
Средний	65,7	[illegible]	[illegible]
Плохой	[illegible]	75,5	[illegible]
Итого	[illegible]	278	453

$$\chi^2(\text{экспериментальное}) = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2(\text{экспериментальное}) = \frac{([illegible])^2}{[illegible]} + \frac{([illegible] - [illegible])^2}{98{,}2} + \frac{(60 - 65{,}7)^2}{65{,}7} + \frac{(1[illegible]0 - 104{,}3)^2}{104{,}3} + \frac{(45 - 47{,}5)^2}{47{,}5} + \frac{(78 - 75{,}5)^2}{75{,}5}$$

$$\chi^2(\text{экспериментальное}) = \frac{([illegible])^2}{61{,}8} + \frac{([illegible])^2}{98{,}2} + \frac{(-5{,}7)^2}{65{,}7} + \frac{(5{,}7)^2}{104{,}3} + \frac{(-2{,}5)^2}{47{,}5} + \frac{(2{,}5)^2}{75{,}5}$$

$$\chi^2(\text{экспериментальное}) = 1{,}09 + 0{,}69 + 0{,}49 + 0{,}31 + 0{,}13 + 0{,}08$$

$$\chi^2(\text{экспериментальное}) = 2{,}79$$

Это экспериментальное (рассчитанное по выборке) значение статистики хи-квадрат следует проверить на значимость.

Шаг 1. Принятие допущений

Модель. Независимые случайные выборки

Зависимая переменная относится к номинальной шкале

Шаг 2. Формулирование нулевой гипотезы

H_0: две исследуемые переменные независимы

(H_1: две исследуемые переменные зависимы).

Шаг 3. Выбор распределения выборочных статистик и задание критической области

Распределение выборочных статистик = распределение хи-квадрат

альфа = 0,05

число степеней свободы = $(r - 1)(c - 1) = (3 - 1)(2 - 1) = 2$

χ^2 (критическое) = 5,991

Шаг 4. Вычисление статистики критерия

$$\chi^2(\text{экспериментальное}) = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2(\text{экспериментальное}) = 2{,}79$$

Шаг 5. Принятие решения

Статистика критерия, χ^2 (экспериментальное) = 2,79, не попадает в критическую [illegible] уровня значимости альфа = 0,05 [illegible] 2 [illegible] χ^2 (критическое) = 5,991. [illegible] не можем отвергнуть [illegible] гипотезу. Наблюдаемые частоты не отличаются статистически значимо [illegible] которые были бы ожидаемы [illegible] если бы переменные [illegible] [illegible] распределялись бы [illegible] случайностью.

[illegible] выборки можно сделать вывод, что академическая успеваемость студентов колледжа не зависит от их семейного положения.

Практический пример 11.1

[illegible] различия в рисовании [illegible] уровнем предвзятости [illegible] [illegible] выборки [illegible]

[illegible] лых и чернокожих американцев были [illegible] отличной от собственной. Полу-
мость к точке зрения или системе убеждений, [illegible]

	Расовая принадлежность		
[illegible]	[illegible]	Чернокожие	Итого
Высокая	15	[illegible]	[illegible]
[illegible]	[illegible]	17	27

[illegible] определить f_e. [illegible]

	Расовая принадлежность		
[illegible]	Белые	Чернокожие	Итого
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Ожидаемые частоты для каждой ячейки находятся по формуле:

[illegible]

$$\frac{(27)(25)}{[illegible]} = [illegible]$$

Значение χ^2 (экспериментальное) равняется:

$$\chi^2 \text{ (экспериментальное)} = \sum [illegible]$$

[illegible]

$$\chi^2 \text{ (экспериментальное)} = 1,79 + 2,03 + [illegible],32 + 1,50$$

$$\chi^2 \text{ (экспериментальное)} = 6,64$$

[illegible] со значением χ^2 (критическое) = 3,841 [illegible] χ^2 (экспериментальное) [illegible]

[illegible] вывод, что существует статистически значимая связь между расовой группой и [illegible]

[illegible] ния США. Мы обнаружили существенное, статистически значимое различие меж-

[illegible]

[illegible] мужчины. Давайте теперь обратимся к исследованию, в котором изучается [illegible]

[illegible] сегрегации по профессиям.

Социологи Хал[illegible] (Hull) и Нельсон (Nelson) опросили случайную выборку из 778 юристов, работающих в Чикаго, об истории их карьеры. Профессия юриста сильно [illegible]

[illegible]

[illegible] распределение мужчин и женщин по перечисленным выше видам юридической практики в [illegible]

Таблица 1. Место работы чикагских юристов в начале карьеры и в настоящее время в зависимости от пола. Частоты и (Проценты*)

	[illegible]	[illegible]	[illegible]	[illegible]
Работающие в большой юридической [illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

в начале своей карьеры женщины: 1) представлены чуть-чуть меньше, чем мужчины [illegible] 2) [illegible] больше представлены [illegible]

Что же обусловило такое поведение? Авторы проверили некоторое количество различных объяснений и пришли к выводу, что имеющиеся доказательства четко указывают на комбинацию половой дискриминации и сознательного выбора женщинами места работы в ответ на эту дискриминацию, а также другие факторы. Женщины все больше [illegible]

[illegible]

11.7. Ограничения на применение критерия независимости хи-квадрат

Как и любой другой критерий значимости [illegible]

[illegible] 5, имея две переменные по 5 категорий каждая, получим таблицу сопряженности 5x5 с 25 ячейками. Такое [illegible] переварить и понять. Есть приблизительное эмпирическое [illegible]

[illegible]

[illegible] процент ячеек таблицы" "большим". Наверное, безопаснее всего предпринять корректирующие действия уже в том случае, когда хотя бы одна из ячеек имеет ожидаемую частоту 5 или менее.

В случае таблиц 2x2 значение χ^2 (экспериментальное) следует скорректировать, применив поправку на непрерывность Йетса (Yates):

Формула 11.4 [illegible]

где χ^2_c — скорректированное значение статистики хи-квадрат;

$|f_o - f_e|$ — абсолютное значение разности между наблюдаемой и ожидаемой частотой данной ячейки.

Поправка заключается в том, что абсолютное значение члена $(f_o - f_e)$ уменьшается на 0,5, и только после этого такая скорректированная разность частот возводится в [illegible]

[illegible]

[illegible] формировать только для того, чтобы удовлетворить условиям применимости статистического критерия. И если вы чувствуете, что категории не удается объединить [illegible]

[illegible] критерий независимости хи-квадрат — при условии, что его результаты будут [illegible]

[illegible]

...ой важностью. С одной стороны, критерии значимости играют в исследовании очень важную роль. И т.к. мы работаем со случайными выборками, мы должны знать, не объясняются ли результаты нашего исследования только одной [illegible]

С другой стороны, как и [illegible]

[illegible]

...рессных методов, [illegible] части [illegible] данной книги.

11.8 Интерпретация статистик: смертная казнь и уровень насильственной преступности

[illegible] течение нескольких последних десятилетий местные, на уровне штата, и федеральные [illegible] власти практически без изменений практикуют жесткий подход к преступности в США. Среди прочих последствий указанного подхода наблюдается увеличение числа смертных приговоров. Применение смертной казни оправдывают тем, что она якобы является устрашающим средством предотвращения насильственных преступлений, т.е. само наличие смертной казни в Уголовном кодексе способствует снижению уровня насильственной преступности. Действительно ли смертная казнь является средством предотвращения тяжелых преступлений? Зависит ли количество убийств или изнасилований от наличия (в законах данного штата) [illegible]

Первым шагом нашего анализа будет построение случайной выборки, состоящей из 25 наблюдений, которые отобраны из списка 50 штатов и округа Колумбия [illegible]

[illegible] мы выпишем уровень [illegible]

[illegible]

...для заданного уровня значимости 0,05 и таблицы 2x2 [illegible] начинается с критического значения 3,841. Мы не можем отвергнуть нулевую гипотезу [illegible] не существует статистически значимой связи между наличием [illegible]

...всех 50 штатах и округа Колумбия [illegible] являются независимыми [illegible] вывод не подтверждает [illegible] о том, что смертная казнь является [illegible]

Таблица 11.7. Уровень убийств в зависимости от наличия смертной казни

Уровень убийства	Наличие смертной казни: [illegible]	Нет	Итого
[illegible]	5	[illegible]	[illegible]
[illegible]	3	12	15
[illegible]	7	8	25

Таблица 11.8. Уровень изнасилований в зависимости от наличия смертной казни

Наличие смертной казни

[illegible]

...политики [illegible] Каков характер [illegible]

[illegible]

результаты применения критерия [illegible] с точки зрения статистики. Но [illegible]

[illegible] взаимосвязи мы бы обнаружили [illegible]

[illegible]

[illegible] смысл. Но мы можем увеличить N, включив в выборку все имеющиеся 5[illegible] [illegible], причем провести проверку для нескольких различных лет в отдельно[illegible]. В профессиональной литературе по социологии встречаются применения критери[illegible] [illegible] будут отражать случайные факторы, отличные от факторов, связанных с генераль[illegible]

[illegible] убийств для трех различных лет, для всех 50-ти штатов и округа Колумбия. Н[illegible] [illegible] отнесено к категории "низкий", если данный уровень убийств меньше ср[illegible]

[illegible] только [illegible]% ([illegible] из 13) штатов без смертной казни [illegible] высокий уровень убийств.

[illegible]

Таблица 11.9. Уровень убийств в зависимости от наличия смертной казни для трех различных лет, для всех штатов и округа Колумбия

A. 1985

Уровень убийств	Наличие смертной казни: Да	Нет	Итого
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

χ^2 (экспериментальное) [illegible]

B. 1991

Уровень убийств	Наличие смертной казни: Да	Нет	Итого
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

χ^2 (экспериментальное) [illegible]

C. 1998

Уровень убийств	Наличие смертной казни: Да	Нет	Итого
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

χ^2 (экспериментальное) 2,35

Таблица 11.10. Уровень изнасилований в зависимости от наличия смертной казни для трех различных лет, для всех штатов и округа Колумбия

A. 1986

Уровень изнасилований	Наличие смертной казни: Да	Нет	Итого
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

χ^2 (экспериментальное) [illegible]

B. 1991

Уровень изнасилований	Наличие смертной казни: Да	Нет	Итого
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

χ^2 (экспериментальное) [illegible]

C. 1998

Уровень изнасилований	Наличие смертной казни: Да	Нет	Итого
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

χ^2 (экспериментальное) [illegible]

[illegible]

Заключение

[illegible]

…ятность того, что это наблюдение будет классифицировано в какую-то [illegible]

2. Так как критерий независимости хи-квадрат является непараметрическим и ему [illegible] …ели можно было удовлетворить. Более того, так для этого критерия [illegible] …и составляет [illegible] — то критерий независимости хи-квадрат [illegible]

[illegible] ожидаемые частоты наблюдений в ячейках, которые бы имели место, если бы переменные были независимы f_e, затем сравнить эти частоты с наблюдаемыми частотами f_o — и так по всем ячейкам. Если нулевая гипотеза справедлива, то ожидаемые и наблюдаемые частоты будут довольно близки между собой. Чем больше различие между наблюдаемыми и ожидаемыми частотами, тем [illegible]

Критерий независимости хи-квадрат имеет несколько серьезных недостатков [illegible] большой размерности таблицы (более 4–5 категорий переменной) становится [illegible] интерпретировать результаты. Кроме того, при уменьшении объема выборки [illegible] …стирования соответствующей статистики. Как и в случае [illegible] объема выборки, применив данный критерий, можно сделать вывод [illegible] …чего статистического критерия, статистическая значимость и [illegible] …теоретической или практической важности.

Сводка формул

[illegible]

Ожидаемые частоты

Число степеней свободы для таблиц сопряженности

11.3 [illegible]

Поправка на непрерывность Йетса (Yates)

11.5
$$\chi_c^2 = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}$$

Словарь

Критерий независимости хи-квадрат. Непараметрический критерий значимости, применимый к переменным, значения которых представлены в виде таблицы сопряженности.

Маргинальные частоты [illegible] в таблице сопряженности.

Наблюдаемая частота (f_o). Частота ячейки таблицы сопряженности, реально наблюдаемая в результате эксперимента.

Независимость. Нулевая гипотеза для критерия независимости хи-квадрат гласит: "Две исследуемые переменные независимы". Две переменные независимы, если для любого наблюдения справедливо: отнесение какого-либо наблюдения в определенную категорию одной переменной не влияет на вероятность того, что это наблюдение будет классифицировано в какую-то определенную категорию второй переменной.

Непараметрический критерий. Критерий, свободный от распределения. Такие критерии не требуют допущения о нормальности распределения, которому подчиняется выборка.

Ожидаемая частота (f_e). [illegible] сопряженности, которая ожидалась бы, если бы переменные были независимы.

Столбец. [illegible]

Строка. Строка таблицы сопряженности. Общепринято, что каждая строка соответствует какой-то [illegible]

Таблица сопряженности. [illegible] распределения двух дискретных [illegible]

Ячейка. Пересечение строки [illegible] ется частота [illegible]

χ^2 (критическое). [illegible] значений хи-квадрат [illegible]

χ^2 (экспериментальное). [illegible] выборки.

Мультимедийные ресурсы

[illegible] Resource Center [illegible] Society (Виртуальное сообщество Центра социологических ресурсов Вадсворта)

[illegible]

[illegible] Statistics: A Tool for Social [illegible] исследования [illegible] широкий спектр ресурсов для изучающих статистику. Начните со щелчка [illegible] чтобы получить доступ к таким материалам

- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]тельные задачи по главам
- [illegible]
- Примеры и упражнения с использованием пакетов статистической обработки данных для MicroCase и SPSS.
- Функциональные диаграммы "Find the test"
- Проверка гипотез для переменных, измеряемых по порядковой шкале.

Задачи

11.1. Для каждой из таблиц а–г рассчитайте χ^2 (экспериментальное). *Подсказка: рассчитайте ожидаемые частоты для каждой ячейки по формуле 11.2. Убедитесь, что вы подставляете в формулу нужные маргинальные частоты по строкам и столбцам! Полезно записать полученные ожидаемые частоты в виде отдельной таблицы — см. табл. 11.4 и 11.3, 11.6 и 11.5. Затем по формуле 11.1 рассчитайте значение статистики критерия. Нужно вычесть ожидаемую частоту из наблюдаемой частоты, возвести эту разность в квадрат, затем разделить на ожидаемую частоту этой же ячейки, — и так по всем ячейкам, а затем [illegible]*

а.

20	25	45
25	20	45
45	45	90

б.

10	15	25
20	30	50
30	45	75

в.

25	15	40
30	30	60
55	45	100

г.

20	45	65
15	20	35
35	65	100

11.2. SOC

[illegible] 25 [illegible] "высокие" или "низкие" по уровню убийств и количеству огнестрельного оружия, проданного в данном городе. Существует ли связь между этими двумя переменными? С помощью одного-двух предложений поясните ваши результаты

	Уровень убийств		
Объем продаж огнестрельного оружия	Низкий	Высокий	Итого
Высокий	8	5	13
Низкий	4	8	12
Итого	12	13	25

11.3. SW

Местный политик озабочена тем, что программа помощи бездомным в ее городе является дискриминирующей по отношению к чернокожим и другим меньшинствам. Данные, приведенные ниже, получены путем опроса случайной выборки чернокожих и белых бездомных. Существует ли статистически значимая связь между расовой принадлежностью и тем, получал ли этот бездомный помощь по данной программе?

	Расовая принадлежность		
Получал ли помощь?	Чернокожие	Белые	Итого
Да	6	[illegible]	[illegible]
Нет	4	8	13
Итого	10	16	26

11.4. PS

Многие аналитики отметили различное поведение мужчин и женщин (gender gap) в выборах президента США, когда женщины чаще, чем мужчины голосовали за кандидата от демократов. Выборку, составленную из преподавателей университета, опросили о том, какую партию они предпочитают. Указывают ли полученные ответы на значимую связь между полом и предпочтением в политике?

	Пол		
Предпочтение в политике	Мужчины	Женщины	Итого
Демократы	10	15	25
Республиканцы	15	0	25
Итого	25	25	50

11.5. PA

Существует ли для государственных служащих связь между уровнем зарплаты и членством в профсоюзе? Данные, приведенные ниже, — это результаты опроса пожарных, из которых сформирована случайная выборка по 100 городам с [illegible] количеством населения. Данные [illegible] разделены на две категории [illegible] значения медианы. Подытожьте [illegible] выводы

Зарплата	Члены профсоюза	Не члены профсоюза	Итого
Высокая	21	29	50
Низкая	14	36	50
Итого	35	65	[illegible]

11.6. [illegible]

[illegible]

11.7. [illegible]

[illegible]

Удовлетворенность	Продолжительность семейной жизни (лет)			
	Менее 5	5–10	Более 10	Итого
Низкая	10	20	[illegible]	[illegible]
Высокая	[illegible]	[illegible]	[illegible]	[illegible]

11.10. PS

Есть ли связь между политической идеологией и классовым (социальным) статусом? Отличаются ли [illegible] студенты, относящиеся к высшему классу, от студентов низшего класса? Связь между этими двумя переменными для случайной выборки объемом 167 студентов колледжа описывается следующей таблицей.

Политическая идеология	[illegible]	[illegible]	Итого
[illegible]	[illegible]	[illegible]	[illegible]

11.11. SOC

В крупном городском колледже примерно половина студентов живет вне территории колледжа в различных [illegible], а другая половина — в общежитиях на территории колледжа. Зависит ли академическая успеваемость от места проживания студента? Ниже представлены результаты для случайной выборки объемом 4300 студентов.

Средний балл (GPA)	[illegible]	[illegible]	[illegible]	Итого
[illegible]	22	33	[illegible]	[illegible]
Умеренный	36	40	[illegible]	[illegible]
Высокий	32	10	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

11.12. SOC

[illegible] выборку мест про[illegible] был оценен по каче[illegible] сие между объективным [illegible]

Удовлетворенность	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

11.13. SOC

Отличается ли уровень поддержки мнения о необходимости легализации марихуаны [illegible] региона? В таблице приведены [illegible] зи между этими двумя переменными для случайной выборки объемом 1020 взрослых граждан. Является ли эта взаимосвязь значимой?

Надо ли легализовать марихуану?	Регион: Север	Средний Запад	Юг	Запад	Итого
Да	60	65	42	78	245
Нет	245	200	180	150	775
Итого	305	265	222	228	1020

11.14. SOC

Исследователь хочет доказать связь между отношением к насилию и агрессивным поведением. Если отношение "вызывает" соответствующее поведение (что очень спорно), тогда люди с положительным отношением к насилию должны иметь более высокие уровни агрессивности поведения. Проведен предварительный тест на 70 респондентах. Среди прочего, у них спрашивали: "Участвовали ли вы в агрессивном инциденте любого рода в течение последних шести месяцев?" Исследователь установил следующую связь.

Участие в инциденте	Отношение к насилию: Благоприятное	Неблагоприятное	Итого
Да	16	19	35
Нет	14	21	35
Итого	30	40	0

По этим данным было рассчитано χ^2 (экспериментальное), равное 0,23, что не является значимым на уровне значимости 0,05 (подтвердите этот вывод [illegible]) [illegible] выборки.

Участие в инциденте	Отношение к насилию: Благоприятное	Неблагоприятное	Итого
Да	1600	1900	1500
Нет	1400	2100	1500
Итого	3000	4000	7000

[illegible] χ^2 (экспериментальное) = 23,4 [illegible] время как предварительный тест не показал значимость? Что же произошло? Считаете ли вы, что второй результат является важным?

11.15. SOC

[illegible] форк, представленной в масштабе США. См. приложение [illegible] полные вопросы и другую информацию об этом опросе. Для каждой таблицы примените критерий независимости хи-квадрат. Для каждой таблицы напишите интерпретацию результата применения критерия (одно-два предложения).

а. Поддержка права на легальный аборт по любой причине в зависимости от возраста

Поддержка абортов	Возраст: Моложе 30	30–49	50 и старше	Итого
Да	154	360	213	727
[illegible]				
[illegible]				

б. Поддержка смертной казни в зависимости от возраста

Поддержка смертной казни	Возраст: Моложе 30	30–49	50 и старше	Итого
За	361	667	675	1003
Против	144	297	262	603
[illegible]				

в. Боязнь гулять в одиночку ночью в зависимости от возраста

Страх	Возраст: Моложе 30	30–49	50 и старше	Итого
Да	147	325	300	772
Нет	202	507	380	077
[illegible]				

Использование SPSS for Windows: критерий независимости хи-квадрат

Демонстрационный пример с использованием SPSS 11.1. Изменяются ли методы воспитания ребенка в зависимости от принадлежности к определенному социальному классу?

В SPSS есть процедура Crosstabs (таблицы сопряженности), которая строит таблицу сопряженности и рассчитывает большое количество статистик. Процедура [illegible] социальных науках, и далее в этой книге вы узнаете немало способов ее [illegible]. Сейчас вы познакомитесь с процедурой Crosstabs, к которой будете обращаться в следующих [illegible].

Изменяются ли методы воспитания ребенка в зависимости от принадлежности к определенному социальному классу? Какой социальный класс более вероятно, чем другие классы, одобрит битье как метод наказания? Мы можем ответить на эти вопросы — по крайней мере, для выборки взятой из данных Объединенного социологического обследования (GSS) за 1998 г., — путем построения таблицы сопряженности, отображающей связь между переменными *class* (класс) и *spanking* (битье) с последующим применением критерия независимости хи-квадрат.

Щелкните Analyze (Анализ), затем Descriptive Statistics (Описательные статистики) и Crosstabs (Таблицы сопряженности). Появится диалоговое окно Crosstabs и его [illegible] для этой переменной гласит FAVOR SPANKING TO DISCIPLINE CHI[illegible] (Одобрение битья как метода наказания ребенка); теперь щелкните по стрелке [illegible] [illegible] Columns [illegible] поле ввода для имени независимой переменной). Щелкните по к[illegible] Statistics (Статистики) [illegible] Chi-square [illegible]

Одобрение битья как метода наказания ребенка — личная классовая идентификация. Таблица сопряженности

		[illegible]				
		Низший класс	Рабочий класс	Средний класс	Высшие слои общества	[illegible]
Битье как метода наказания ребенка	Согласен	4	110	100	11	
	Не согласен	20	204	179	1	
	Категорически не согласен	8	65	84	1	

Критерий хи-квадрат

	Статистика	Ст. св.	Значимость (двусторонняя)
[illegible] хи-квадрат Пирсона	[illegible]	[illegible]	
[illegible]	[illegible]	9	0,22[illegible]
Критерий хи-квадрат с поправкой Мантеля-Хэнзеля (мера линейной зависимости между строками и столбцами)	3,068	1	0,080
Число валидных наблюдений	880		

В таблице сопряженности, как обычно, представлены наблюдаемые частоты — [illegible] критерия независимости хи-квадрат идут вслед за таблицей сопряженности. Мы видим, что χ^2 (экспериментальное) = 11,486, число степеней свободы (ст. св.) равно [illegible], а значимость (точная вероятность получения такого результата при условии справедливости нашей нулевой гипотезы) равна 0,244. Это намного больше уровня значимости альфа = 0,05, которым обычно задаются, поэтому мы можем сделать вывод [illegible] отсутствии статистически значимой связи между переменными *class* (класс) и *spanking* (битье). Итак, одобрение битья как метода наказания ребенка не зависит от [illegible] социального класса.

Демонстрационный пример с использованием SPSS 11.2. Кто смотрит порнофильмы?

[illegible] Crosstabs [illegible] это поле ввода для имени зависимой переменной, а Columns — это поле ввода для имени независимой переменной). Не забудьте в дополнительном окне заказать хи-квадрат [illegible]

Смотрел ли респондент хоть один порнофильм в прошлом году — личная классовая идентификация. Таблица сопряженности

		Личная классовая идентификация				
		Низший класс	Рабочий класс	Средний класс	Высшие слои общества	Итого
[illegible]	[illegible]	[illegible]	[illegible]	97	4	224
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Критерий хи-квадрат

	Статистика	Ст. св.	Значимость (двусторонняя)
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		2
Критерий хи-квадрат с поправкой [illegible]			
[illegible]	3		
Число валидных наблюдений	923		

[illegible] наблюдаемыми переменными нет значимой связи. Точная [illegible]

Демонстрационный пример с использованием SPSS 11.3. Блаженны ли незнающие?

Мы можем проверить справедливость древнего афоризма "блаженны незнающие" (ignorance is bliss), для чего воспользуемся данными Общенационального социологического обследования (GSS) 1998 г. Являются ли более образованные люди менее счастливыми? Чтобы проверить это, выберем переменную *degree* в качестве меры образования, а в качестве меры счастья — переменную *happy* (*Happy* — это ответ респондента на вопрос GSS, в котором его просили оценить уровень общего счастья [illegible] — т.к. в GSS есть еще, например, вопрос о счастье в браке и др.). В [illegible] кодировке GSS, *degree* имеет 5 категорий (см. приложение Ж). Но для [illegible] мы сведем число категорий к двум: 1) люди с образованием только [illegible] меньшим, 2) люди с образованием, большим, чем школа. Выполните [illegible] переменной *degree* с помощью команды Recode into Different Variable [illegible] другое имя (например, *rdeg*). Введите в боксе Old [illegible] Recode into Different Variable [illegible] ровать в другую переменную) указания по перекодировке категорий:

[illegible]

Вызовите процедуру Crosstabs [illegible] мионный пример 11.[illegible]. Укажите [illegible] *happy* [illegible] этой переменной GENE[illegible] мой переменной — *rdeg* [illegible] [illegible] для оценки независимой переменной [illegible] расчет [illegible]

Общее счастье * RDEG Таблица сопряженности

		RDEG		Итого
		1.00	2.00	
Общее счастье	Очень счастлив	[illegible]	[illegible]	[illegible]
	Довольно счастлив	[illegible]	[illegible]	[illegible]
	Не слишком счастлив	121	29	[illegible]
Итого		[illegible]	[illegible]	[illegible]

Критерий хи-квадрат

	Статистика	Ст. св.	Значимость (двусторонняя)
Критерий хи-квадрат Пирсона	[illegible]	2	[illegible]
Отношение правдоподобия	20.[illegible]	2	[illegible]
Критерий хи-квадрат с поправкой Мантеля-Хензеля / Линейная связь между строками и столбцами	8.691	1	[illegible]
Число валидных наблюдений	[illegible]		

В таблице сопряженности, как обычно, представлены наблюдаемые частоты для каждой ячейки и маргинальные по строкам и столбцам, а также указано общее число наблюдений. Результаты применения критерия независимости хи-квадрат идут вслед за таблицей сопряженности. Мы видим, что χ^2 (экспериментальное) = 19.730, число степеней свободы равно 2, а значимость (точная вероятность получения таких результатов при условии справедливости нашей нулевой гипотезы) равна 0,000. Так как значимость = 0,000 (что заведомо меньше уровня значимости альфа = 0,05, которым обычно задаются), то мы отвергаем нулевую гипотезу о независимости переменных. [illegible]

[illegible] нительный анализ. Если никаких [illegible] образованных людей должен быть счастлив [illegible] таблица сопряженности показывает прямо [illegible] "счастливым" относят себя 29,94% (276 из 922) [illegible] 38,84% (174 из 448) более образованных [illegible]

Упражнения

11.1. Как продолжение демонстрационного [illegible] Crosstabs [illegible] Cross tabulation [illegible]

[illegible] (Помните, что в таблице сопряж[illegible]

[illegible]

11.2 [illegible]

наиболее значимы?

11.3. Так как мы уже перекодировали переменную *degree* (см. демонстрац[illegible] пример 11.3), давайте выясним, имеет ли образование [illegible] связь с *grass*, *gunlaw* или *cappun*. Запустите процедуру Crosstabs (в SPSS) [illegible] Cross-tabs [illegible] (в Microcase), указывая в качестве зависимых поочерёдно [illegible] переменные (*grass*, *gunlaw* и *cappun*), а в качестве независимой переменной [illegible] (перекодированную *degree*). (Помните, что в таблице сопряженности [illegible]

[illegible] выводы [illegible] абзаца). Какие взаимосвязи оказались [illegible]

Итоговые упражнения к части II

1. Выберите критерий значимости, наиболее подходящий для каждой ситуац[illegible] [illegible] (например, по возрастанию сложности [illegible] часть I).

а. Существует ли различие между мужчинами и женщинами (gender gap), касающееся пользования Интернет? Среди мужчин (случайная выборка 1) и женщин (случайная выборка 2) проведен опрос на тему: "Сколько минут в неделю в среднем Вы проводите в Интернете, независимо от цели?" Является ли это различие значимым?

Женщины	Мужчины
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

б. Существует ли связь между социальным классом и вовлеченностью школьников в такие виды деятельности, как клубы и спорт? Собраны данные для случайной выборки школьников. Является ли эта связь значимой?

	Социальный класс		
Вовлеченность	Средний	Низший	Итого
Высокая	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

в. [illegible] от уровня образования?

Ниже школьного	Только школа	По крайней мере колледж
[illegible]	[illegible]	[illegible]

г. [illegible]

[illegible] 3,5

[illegible] = 1450

[illegible]

2. Ниже приведен список вопросов (заданий), на которые можно ответить с помощью статистических методов, описанных в главах 7–11. Для каждого задания следует выбрать наиболее подходящий критерий значимости, рассчитать необходимые статистики, сформулировать ваши выводы.

Некоторые из заданий таковы, что вам вначале придется вычислить выборочные статистики (например, средние или проценты), которые нужны для критерия значимости. Везде задавайтесь уровнем значимости альфа = 0,05. Задания не упорядочены каким-то определенным образом (например, по возрастанию глав, входящих в часть II). Для каждой главы есть по крайней мере одно задание, но я не [illegible] придумать задание для каждой статистической процедуры, описанной [illegible]

[illegible] выбрать наиболее подходящий критерий значимости и процедуру, вы [illegible] во внимание смысл вопроса, количество сравниваемых выборок [illegible], а также шкалу измерений переменных. Это упражнение основано на [illegible] Общенационального социологического обследования (GSS). (См. [illegible] где вы найдете полные вопросы и схемы кодирования.) [illegible] я подготовил, будут сокращенные коды (см. [illegible]). Значения некоторых переменных GSS упрощены. [illegible]

а. [illegible]

б. [illegible]

в. Оцените среднее количество часов, [illegible] TV [illegible]

г. Считается, что в среднем американцы имеют по 2,3 ребенка. [illegible] борка, приведенная в таблице в конце [illegible]

д. Различается ли значимо образованность католиков и протестантов?

е. Изменяется ли среднее количество часов, проводимых за просмотром TV, в зависимости от уровня счастья?

ж. На основе данных выборки, приведенной в таблице в конце этого упражнения, оцените процент черных американцев в населении США.

ИССЛЕДУЕМЫЕ ПЕРЕМЕННЫЕ

1. Сколько у вас детей — в течение всей жизни (CHILDS). Значения этой переменной не кодированы.

2. Уровень образования респондента (DEGREE)

   0 — Меньше школьного

   1 — Только школа

   2 — По крайней мере колледж

3. Раса (RACE)

   1 — Белые

   2 — Черные

4. Возраст (AGE)

   1 — Моложе 35

   2 — 35 и старше

5. Количество часов в день, проводимых за просмотром TV (TVHOURS). Значения этой переменной не кодированы.

6. Ваши религиозные предпочтения [illegible]

   1 — Протестант

   2 — Католик

7. Доход респондента (INCOME91)

   1 — 24999 или менее

   2 — 25000 или более

8. Уровень общего счастья респондента (HAPPY)

   1 — Очень счастлив

   2 — Довольно счастлив

   3 — Не слишком счастлив

Номер наблюдения	Число детей	Уровень образования	Раса	Возраст	Количество часов в день, проводимых за просмотром TV	Религиозные предпочтения	Доход	Счастье
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Часть III

МЕРЫ СВЯЗИ

12 Связь между двумя переменными. Введение и основные понятия

Цели изучения

К концу этой главы вы сможете

1. Объяснить, как можно использовать меры связи для описания и анализа [illegible] между двумя переменными [illegible] в отличие от ее значимости.
2. Определять связь в контексте двухвходовых таблиц сопряженности и в терминах изменяющихся условных распределений.
3. Приводить и объяснять три характеристики парной связи (существование, силу, структуру или направление).
4. Исследовать связь между двумя переменными путем [illegible] с вычисления процентных значений для таблицы [illegible]

12.1 Статистическая значимость и теоретическая важность

[illegible]

Начиная с главы 8 [illegible]

[illegible]

[illegible] переменная X является причиной изменения переменной Y [illegible]

Итак, меры связи [illegible]

Невзирая на их полезность, меры связи все же имеют свои ограничения. [illegible]

[illegible] связи между двумя переменными. [illegible]

[illegible] Детальнее причинно-следственные связи будут рассмотрены в IV части, [illegible]

[illegible]

12.2 Связь между переменными и таблица сопряженности

Как правило, [illegible]

производстве интересует связь между удовлетворенностью работой на сборочном конвейере своей работой и производительностью их труда. Если между этими двумя [illegible] производительности [illegible]

[illegible]

их коллеги с низкой удовлетворенностью. Следовательно, уровень производительности труда меняется в зависимости от уровня удовлетворенности.

Указанная связь становится более очевидной при использовании **таблиц сопряженности**. Как говорилось в главе [illegible], таблицы сопряженности являются [illegible] представлением частот наблюдения определенных значений [illegible] двух различных переменных. Условно принято размещать **независимую переменную** [illegible] и **переменную** [illegible], то есть переменную, принимающую случайные значения, в столбцах таблицы, а зависимую переменную [illegible] переменную [illegible] [illegible] бце таблицы (по вертикали) представлены числовые величины или категории независимой переменной X, а в каждой строке (по горизонтали) — числовые величины или категория зависимой переменной Y.

[illegible]

ностью работой для вымышленной выборки, состоящей из 173 заводских рабочих. Чтобы обнаружить наличие зависимости между переменными таблицы, рассмотрим столбцы. В каждом столбце приведено распределение частот значений зависимой переменной [illegible] независимой переменной. Например, при рассмотрении [illegible]

[illegible] и [illegible] высокую. Средний столбец показывает, что [illegible] из 61 рабочего со средней удов[illegible]

[illegible]

Таблица 12.1. Зависимость производительности труда от удовлетворенности своей работой ([illegible] величины)

Производительность труда (Y)	Удовлетворенность своей работой (X): Низкая	Средняя	Высокая	Всего
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
В[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Н[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible] бующие" частотные [illegible]
Y, поскольку отра[illegible] условными распределениями переменной [illegible]

[illegible]

[illegible]

[illegible] ностью труда и удовлетворенностью своей работой [illegible] распределение частот значений [illegible]

[illegible] рабочих с низкой удовлетворенностью своей работой также [illegible] низкую производительность труда. С другой стороны, высокую производительность имели более [illegible] рабочих с высокой удовлетворенностью.

[illegible] хи-квадрат, [illegible] для проверки значимости [illegible] может быть использован для обнаружения зависимости между двумя переменными [illegible] в табличном формате. Любое значимое значение статистики [illegible] на шаге 4, указывает на существование зависимости. Например, [illegible] таблицы 12.1 [illegible] значение критерия хи-квадрат равно 24,2. Оно [illegible] [illegible] сделанный на основе условных распределений [illegible] о [illegible] между удовлетворенностью своей работой и производите[illegible]

[illegible]

Часто бывает, что к моменту рассмотрения вопроса о существовании связи [illegible] с использованием критерия хи-квадрат уже выполнен. В таких случаях [illegible] углубленного изучения условных распределений Y, чтобы удостовериться в существовании [illegible] двумя переменными уже нет необходимости. Если [illegible] хи-квадрат равен нулю, то рассматриваемые [illegible] [illegible] любое значимое отличие значения хи-квадрат свидетельствует о существовании связи между переменными. Однако не следует забывать о том, что статистическая значимость и взаимосвязь — это разные понятия [illegible] случай, когда существует взаимосвязь между двумя переменными [illegible] [illegible]

[illegible]

[illegible] связь. В следующем разделе изучение зависимости уже не будет [illegible] Систематическим образом будет [illegible] показано, как может быть [illegible]

12.3. Три характеристики парных статистических связей

[illegible]

ответов на следующие вопросы.

1. Существует ли связь?

[illegible]

3. Каков характер этой связи?

Ниже эти вопросы будут рассмотрены по [illegible]

Существует ли связь?

[illegible]

абсолютных величин рассчитанными процентами (примечание: в новой таблице не [illegible] ния в скобках). Таблицы с процентными величинами, кроме того, что позволяют [illegible]

Таблица 12.2. Зависимость производительности труда от удовлетворенности своей работой (процентные величины)

Производительность труда (Y)	Удовлетворенность своей работой (X)		
	Низкая	Средняя	Высокая
Низкая	50,00%	34,43%	13,46%
Средняя	33,33%	[illegible]	[illegible]
Высокая	16,67%	24,59%	51,92%
Всего	[illegible]	[illegible]	[illegible]

Теперь, когда мы [illegible] наибольшим числом на [illegible] [наибольш]ая ячейка) в каждом столбце занимает другое место. Для рабо[тников] [illegible] [удов]летворенностью своей работой наибольшая ячейка находится [illegible] [ни]зкая производительность труда). В среднем [illegible] [наи]большая ячейка находится в средней строке [illegible] находится в нижней строке [illegible] в правом столбце (высокая удовлетворенность) [illegible] переменными действительно существ[ует] [illegible]

[illegible] волос и производительностью труда. И это всего лишь единичный пример из множества аналогичных, указывающих на отсутствие связи. Главный признак — это неизменность условных распределений Y. Уровень производительности труда вообще не изменяется при различном цвете волос. Следовательно, между этими двумя переменными полностью отсутствует статистическая связь. К тому же, если рассчитать для данной таблицы хи-квадрат, его значение окажется равным нулю, что также является признаком отсутствия связи.

Таблица 12.3. Зависимость производительности труда от цвета волос (иллюстрация полного отсутствия статистической связи)

Производительность труда (Y)	Цвет волос (X)		
	Каштановые	Белокурые	Рыжие
Низкая	[illegible]	[illegible]	[illegible]
Средняя	[illegible]	[illegible]	[illegible]
Высокая	[illegible]	[illegible]	[illegible]
	100%	[illegible]	[illegible]

Насколько сильна связь?[7]

После того, как установлено существование связи, требуется каким-то образом оценить, насколько она сильна. По существу — это вопрос определения величины [illegible] отсутствие изменений переменной Y (см. таблицу 12.4).

Таблица 12.4. Зависимость производительности труда от цвета волос (иллюстрация полной зависимости)

Производительность труда (Y)	Цвет волос (X)		
	Каштановые	Белокурые	Рыжие
Низкая	[illegible]	[illegible]	[illegible]
Средняя	0%	[illegible]	0%
Высокая	100%	[illegible]	100%

[illegible]

Полная связь должна рассматриваться [illegible] ... [illegible] отсутствием связи, по крайней мере [illegible]

[illegible]

[illegible] ной. К примеру, мы, что у некоторого рабочего квалификацио[illegible] [illegible]

[illegible]

Но, конечно, подавляющее большинство существующих связей [illegible] лежат между двумя крайними значениями — случаями полного отсутствия и [illegible]

[illegible] ...тельный и действенный способ определения этих промежуточных [illegible] связи. Например, таблицы 12.1 и 12.2 свидетельствуют о существовании статистической связи между производительностью труда и удовлетворенностью [illegible]

Но каким образом можно количественно выразить силу или [illegible] насколько она близка к полной? Или насколько она отличается от слу[illegible]

[illegible] ...мотря на возможность образного описания связи, задача [illegible] в случае наличия у нас некоторой системы показателей [illegible] ...тельной оценки тесноты связи (например, сильная, средняя [illegible]). Как будет показано в главах 13–15, эти объективные показатели тесноты связи основываются на различных мерах связи. Фактически все они построены так [illegible]

[illegible] 1,00 (или +1,00 для порядковых (ординальных) и интервально-относительных мер связи). Значение, равное 0,00, указывает на полное отсутствие связи между переменными (изменение условного распределения Y отсутствует), а 1,00 (или +1,00 для порядковых и интервально-относительных мер связи) говорит о существовании [illegible] [illegible]

[illegible] ...ерждение: чем ближе значение к 1,00, тем теснее связь (т.е. тем больше [illegible] условных распределений Y).

Практический пример 12.1

[illegible] ... кинозала пережить (с п[illegible] ... подобных лент [illegible]

[illegible] ... В [illegible] ...ные распределения переменных "страх" и "насилие" в процентном выраже[illegible] ... (600) [illegible]

Частота посещений кинотеатра	Страх: Слабый	Умеренный	Сильный
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
Часто	20	20	40
	[illegible]%	100%	100%
	(200)	(200)	(200)

Условные распределения частоты посещений (Y), действительно изменяются при различных уровнях страха (X), что доказывает существование между этими двумя переменными статистической связи. Группировка частот по диагонали от верхнего левого к нижнему правому углу предполагает наличие существенной связи в предсказанном направлении. Люди, слабо подверженные страху в своей ежедневной жизни, редко ходят на просмотры фильмов с насилием, и напротив, часто посещают кинотеатр те, кто испытывает сильные страхи. Полученные результаты означают существование тесной связи между страхом и частотой посещений. Из их рассмотрения также можно сделать вывод о наличии интересной причинной проблемы. Данные таблицы поддерживают мысль о том, что напуганные люди обращаются к фильмам со сценами насилия как к механизму по преодолению своих страхов (X является причиной Y). В то же время таблица предлагает доказательства противоположного причинного аргумента: посещение фильмов со сценами насилия увеличивает опасения человека за свою личную безопасность (Y является причиной X). То, что результаты исследования подтверждают оба причинных аргумента, еще раз свидетельствует о различии понятий статистическая и причинно-следственная связь.

[illegible]

Предложенные в данной книге условные правила построения и интерпретации таблиц [illegible] сопряженности являются широко, однако не повсеместно, распространенными [illegible] ... чинами заполняют столбцы. Однако [illegible] ...женности, следует проверить [illegible] ...данными условными правилами. [illegible] строки, а зависимой — столбцы [illegible] перерисовать таблицу [illegible] обратить на то, в правильном [illegible] Даже опытные профессионалы [illegible] [illegible]

[illegible]

Статистики в профессиональной литературе

Исследовательницы Мишель Келли (Michelle Kelly) и Бет Парсонс (Beth Parsons) прове- [illegible] распространенность в стенах этого учебного заведения сексуальных домогательств. Их интересовало, насколько новое федеральное законодательство, получившие известность в национальном масштабе сексуальные скандалы и изменение политики университета в этом вопросе повлияли на уровень домогательств к женщинам в отдельном учебном заведении (включая студенческий городок). В выборку из 765 респондентов были вклю- [illegible] преподаватели [illegible] административные работники), частично студентки и аспирантки. Студенток отбирали из [illegible] из общеуниверситетского списка аспирантуры, т.е. независимо от специализации. Была использована следующая градация сексуальных домогательств: оскорбление по половому признаку (словесные оскорбления в сексуальном контексте, враждебные или уничижительные высказывания или отношение), нежелательное внимание со стороны лиц [illegible]

Келли и Парсонс особое внимание уделили оценке общего масштаба проблемы, [illegible] личиям по виду сексуальных домогательств и [illegible] обнаружено, что 30% обслуживающего персонала, 22% препода- [illegible] отношению к ним. Для сравнения [illegible] довании 20–40% студенток и [illegible] Различия по виду сексуальных [illegible] сексуальные домогательства

Таблица 1. Сексуальные домогательства в зависимости от статуса респондента

Вид сексуального домогательства	Сотрудница	Студентка
Оскорбление по половому признаку	[illegible]	[illegible]
[illegible] со стороны лиц мужского пола	[illegible]	[illegible]
Принуждение к близости	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

Примечание. [illegible] *сексуальным домогательствам по отношению к ним сообщили 158 сотрудниц и 87 студенток.*

Из таблицы можно сделать два очевидных вывода: сотрудницам [illegible] сексуальными [illegible] нежелательное внимание [illegible] из таблицы? Критерий хи-квадрат показывает статистическую значимость связи между переменными (χ^2 = [illegible], p < 0,01), но в то же время результат может быть не очень существенным, поскольку в выборку включены все сотрудницы и отобранные неслучайным образом студентки. Случайным образом были выбраны только аспирантки.

Келли и Парсонс также выяснили, что женщины в их выборке чаще реагировали на сексуальные домогательства уклончиво (64%), а не решительно (36%). Другими словами, женщины более склонны проигнорировать непристойное поведение мужчин или просто поговорить о происшедшем с подругой, чем пожаловаться на обидчика какому-нибудь официальному лицу. Также было обнаружено, что реакция жертвы домогательства очень зависит от того положения, которое занимает в университете оскорбивший ее мужчина (см. таблицу 2). Домогательства игнорировались вдвое чаще, если обидчиком выступал студент, а не сотрудник университета, тогда как более решительные действия (устная жалоба вышестоящему начальнику обидчика или даже подача письменной жалобы) более вероятны в случае домогательств со стороны сотрудника.

Таблица 2. Реакция на сексуальные домогательства в зависимости от статуса обидчика

Реакция	Статус обидчика: Сотрудник	Студент
[illegible]	[illegible]	[illegible]
Разговор об инциденте с подругой	20%	22%
[illegible]	[illegible]	[illegible]
Просьба к мужчине прекратить его действия	17%	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

[illegible] Келли и Парсонс [illegible] что женщины [illegible] реагировать на сексуальные домогательства [illegible] образом [illegible] изменений в общественном [illegible] университета негативное отношение к сексуальным домогательствам и даже [illegible] жесткую политику в этом плане, принятую университетом [illegible] дования Келли и Парсонс, женщины [illegible] предпочтение неформальным и некардинальным мерам противодействия [illegible] Авторы предполагают, что [illegible] политика в вопросе сексуальных домогательств не является [illegible] действий и для побуждения [illegible]

Источник: выдержки из [illegible] Higher Education [illegible]

[illegible] Однако помните, что статистическая значимость и [illegible] понятия, поэтому хи-квадрат сам по себе не [illegible] любое отличие от нуля значение этого критерия [illegible]

[illegible] хи-квадрат может [illegible] Тем не менее, существуют [illegible] статистические показатели [illegible] между переменными. Эти способы [illegible]

[illegible] связи [illegible] в конце настоящей главы.

Какой характер [illegible] или направление [illegible]

[illegible] требуется [illegible]

[illegible] между сравнительной сложностью труда и удовлетворенностью своей работой. Данные таблицы 12.2 свидетельствуют о том, что [illegible] показатель удовлетворенности связан с [illegible] сти: средняя удовлетворенность [illegible] высокая удовлетворенность — с высокой [illegible]

Если обе переменные относятся к [illegible] характер связи может иметь [illegible] ной, если переменные изменяются в одном направлении [illegible] положительной [illegible]

[illegible] случае положительной связи при увеличении значения [illegible] ной следует за уменьшением первой. Таблица 12.5 (составлена по [illegible]

[illegible] удовлетворенностью своей работой в рассмотренном примере (см. таблицы 12.1 и 12.2).

Таблица 12.5. Зависимость частоты [illegible] от уровня [illegible] (иллюстрация положительной связи)

Частота посещения библиотек	Уровень образования		
	Низкий	[illegible]	[illegible]
Средняя	[illegible]	20%	[illegible]
Высокая	10	[illegible]	[illegible]

При отрицательной связи переменные изменяются в противоположных направлениях. Большие числовые значения одной переменной связаны с меньшими значениями другой, и увеличение одной из них сопровождается уменьшением другой. [illegible] том образования время телепросмотров сокращается.

Таблица 12.6. Зависимость количества времени телепросмотров от уровня образования (иллюстрация отрицательной связи)

Количество времени телепросмотров	Уровень образования		
	Низкий	Средний	Высокий
Малое	[illegible]	[illegible]	[illegible]
Среднее	30	60	30
Большое	60	20	10
100%	100%	100%	[illegible]

Меры связи для порядковых и интервальных/относительных переменных построены таким образом, что в случае положительной связи они принимают положительные значения, а в случае отрицательной связи — отрицательные. Следовательно, значение меры связи со знаком плюс указывает на положительную связь между переменными, причем значение +1,00 говорит об полной положительной связи. Знак [illegible] является признаком отрицательной связи, об абсолютности которой свидетельствует значение −1,00. *Для практического освоения методики определения характера связи читатели могут рассмотреть любую из задач, приведенных в конце настоящей главы. Процедура определения направления связи освещена в задачах 12.1, 12.6, 12.7, 12.8, 12.9, 12.10 и 12.12.*

Заключение

1. Анализ связи между переменными дает информацию, дополняющую [illegible] для обнаружения неслучайных взаимосвязей, тогда как меры связи служат количественной оценкой важности или силы связи.
2. [illegible] характеристики существование статистической связи [illegible] характеристики могут быть исследованы путем вычисления процентов для таблицы сопряженности в направлении независимой переменной [illegible] тикали) с последующим их [illegible]
3. Таблицы 12.1 и 12.2 могут [illegible] характеристик. Очевидно [illegible] тельностью труда [illegible] деления зависимой пере[illegible]

[illegible] рабочие с низкой удовлетворенностью работой [illegible]

[illegible] труда ("удовлетворенных рабочих и работа спорится"). Эти результаты [illegible] переменными, но в то же время сами по себе они не могут служить доказательством [illegible] как статистическая связь и причинно-следственная связь — разные понятия. В данном примере, хотя в качестве независимой переменной рассматривалась удовлетворенность работой, можно было бы за независимую переменную принять производительность труда, т.е. рассматривать обратную причинную связь ("спорится работа — значит, довольны рабочие"). Так, анализ таблиц 12.1 и 12.2 [illegible]

Анализ связи между двумя переменными может рассматриваться как очень важный процесс, в ходе которого систематично собираются свидетельства в пользу (или [illegible]) [illegible] (на практике доказать существование причинно-следственной связи достаточно трудно). Как будет показано в IV части, для обнаружения возможных причинно-следственных связей в одних могут быть использованы некоторые многомерные методы.

Словарь

Зависимая переменная. [illegible] переменная, рассматриваемая как эффект (следствие).

Мера связи. Статистические показатели, [illegible]

Независимая переменная. При взаимосвязи между [illegible] рассматриваемая [illegible]

Отрицательная связь. Связь между двумя [illegible] изменяются в противоположном [illegible]

[illegible] корреляция (зависимость, связь). Связь между двумя [illegible], при которой переменные изменяются в [illegible] переменная возрастает, то возрастает и другая, а [illegible] другой переменной.

Статистическая связь. Взаимосвязь между двумя (или более) переменными. Говорят, что между двумя переменными существует связь, если распределение одной переменной изменяется при различных значениях или частотах другой переменной.

Условное распределение Y. Распределение численных значений зависимой переменной при определенном значении или частоте независимой переменной в случае представления этих двух переменных в табличной форме.

X. Символ, используемый для обозначения любой независимой переменной.

Y. Символ, используемый для обозначения любой зависимой переменной.

Мультимедийные ресурсы

[illegible]

На web-сайте компании вы сможете найти шестое издание книги *Statistics: A Tool for Social Research* (Статистика как инструмент социологических исследований), а также получите доступ к разнообразным ресурсам, полезным для тех, кто изучает социологию. Зайдя в раздел Student Resources (Ресурсы для студентов) на web-странице книги, вы найдете следующие материалы.

- Обзор основ математики
- Обзор статистики
- Обучающие карты,
- Интернет-ссылки
- [illegible]
- Таблицу случайных чисел
- Примеры и упражнения с использованием пакетов статистической обработки данных MicroCase и SPSS
- Функциональные диаграммы "Find the test"
- Проверку гипотез для переменных, измеряемых по порядковой шкале

Задачи

12.1 Ра[illegible]

[illegible] ность работы отдела. Существует ли [illegible] между переменными? Какова теснота [illegible]

[illegible]	Уровень авторитаризма	
	Низкий	Высокий
[illegible]	[illegible]	[illegible]

[illegible]

[illegible] [illegible] платы за обучение и полом, видом дисциплины, а также видом контракта, по которому работает данный преподаватель. Охарактеризовать силу и направление связи.

а. Поддержка увеличения платы за обучение в зависимости от пола преподавателя.

Поддержка	Мужчины	Женщины
За	12	8
Против	15	12
Всего	27	20

б. Поддержка увеличения платы за обучение в зависимости от вида преподаваемой дисциплины.

Поддержка	Гуманитарные науки	Естественные науки и бизнес
За	6	13
Против	14	14
Всего	20	27

в. Поддержка увеличения платы за обучение в зависимости от вида контракта, по которому работает данный преподаватель.

Поддержка	Одногодичный контракт	Многолетний контракт
[illegible]	15	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

12.3. PS

[illegible]

	Выборы 1996 года	
Выборы 2000 года	Демократ	Республиканец
[illegible]	117	23
[illegible]	[illegible]	[illegible]
Всего	136	201

12.4. SOC

Среди большой группы пенсионеров был проведен опрос с целью выяснить, насколько им нравятся танцевальные вечера отдыха. Следует дать ответ на вопрос: существует ли связь между полом пенсионера и потребностью в вечерах отдыха? Также требуется в нескольких предложениях описать силу и характер этой связи.

Устраивать ли больше вечеров отдыха?	Мужчины	Женщины
Да	321	426
[illegible]	175	251
[illegible]	496	677

12.5. В задачах 11.[illegible] и 11.15 (глава 11) рассчитать проценты по величинам в направлении независимой переменной. Имеется ли статистическая связь между [illegible]

12.6. SW

Как главный психиатр штата вы заметили, что в некоторых лечебных учреждениях психиатрического профиля очень высока текучесть кадров среди [illegible] [illegible] ется низким уровнем образования в администрировании и плохо развитыми [illegible] ной на решения проблемы, вы [illegible]

Существует ли связь между текучестью [illegible] директоров? Описать силу и характер этой связи.

[illegible]	Наличие [illegible] у директоров	
	Да	Нет
[illegible]	[illegible]	[illegible]
Средний	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

12.7. CJ

В крупном городе [illegible]

	Начинали употреблять	
Уровень преступности	**Да**	**Нет**
Низкий	29	15
Средний	33	27
Высокий	52	45
Всего	114	87

12.8. PS

[illegible] заявили, что они придерживаются [illegible] [illegible] ее о легализации марихуаны. Требуется ответить, существует ли статистическая связь между этими двумя переменными. В нескольких предложениях опи[illegible]

Отношение к легализации марихуаны	Либералы	Умеренные	Консерваторы
[illegible]	132	78	52
[illegible]	[illegible]	[illegible]	[illegible]
Всего	233	165	161

12.9. SOC

Исследователь выдвинул гипотезу, что внешняя привлекательность связана с академической успеваемостью (т.е., чем более привлекательна студентка, тем более высокие оценки она получает). В ходе исследования 100 студенткам был присвоен рейтинг внешней привлекательности. Одновременно путем опроса были собраны сведения о их средних отметках. На основе полученных данных была построена таблица сопряженности. Требуется рассчитать процентные величины для этой таблицы в правильном направлении и сравнить условные распределения. Существует ли связь между переменными? Какова, по вашему мнению, сила этой связи? Каково ее направление? Подтверждают или нет полученные фактические данные гипотезу исследователя? И [illegible] чему?

	Внешняя привлекательность		
Средние оценки	Низкая	Умеренная	Высокая
[illegible]	[illegible]	6	5
[illegible]	[illegible]	7	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

12.10. SOC

[illegible]

[illegible] некоторого исследователя заинтересовал вопрос о том, насколько степень распространенности этих чудачеств зависит от их освещения в местной университетской прессе. Возможно, статьи о новом "приколе" лишь разжигают интерес к нему и подталкивают студентов к тому, чтобы тоже совершить модное в данный момент чудачество. Для выборки из 25 университетов были присвоены оценки степени освещения чудачеств в местной прессе (как "интенсивное", "умеренное" и "отсутствие") и уровню распространенности среди студентов этого нового модного увлечения. Информация по университетам приведена ниже. Требуется представить эти данные в процентной форме в таблице с соответствующими названиями столбцов и строк. Необходимо ответить на вопрос: свидетельствует ли данная таблица о существовании связи между освещением в прессе и распространенностью нового чудачества?

Университет	Освещение в прессе	Распространенность увлечения
1	Интенсивное	Широкая
2	Интенсивное	Умеренная
3	Умеренное	Умеренная
4	Умеренное	Умеренная
5	Умеренное	Широкая
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	Интенсивное	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

[illegible]

Использование SPSS for [illegible] связи между двумя переменными

[illegible] и к просмотру кинофильмов для [illegible]

[illegible] по Analyze. Descriptive Statistics (А[illegible]) [illegible] ы сопряженности) [illegible] общественный класс [illegible]

[illegible] Continue (Продолжить) и OK при этом на экран будет выведена следую[illegible]

Зависимость веры в загробную жизнь от принадлежности к общественному классу

[illegible]

			Принадлежность к общественному классу (согласно самоидентификации респондента)				
			[illegible]	[illegible]	[illegible]	Высший класс	Всего
Вера в [illegible]	Да	Количество	30	370	374	25	799
		% по столбцу	65,2%	81,7%	82,4%	73,5%	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Всего		Количество	46	453	454	[illegible]	[illegible]
		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Критерий хи-квадрат

	Значение	df (степени свободы)	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Количество действительных случаев	[illegible]			

[illegible] вывод по какой причине [illegible] распределения [illegible] по всех [illegible]

Рассмотрим верхнюю строку таблицы (Yes). Можно видеть приблизительно 15- [illegible] высшего класса, даже снижается приблизительно на 10%. Однако в случае самой [illegible]

Что касается направления связи, то она, в общем, положительна: при переходе от более низкого класса к более высокому вера в загробную жизнь (% "yes") тоже возрастает. Однако отметим, что возрастание процентных величин не постоянно, поскольку имеется снижение процента верящих в загробную жизнь при переходе от среднего класса к высшему. Поскольку значение критерия хи-квадрат равно 9,355 при уровне значимости менее 0,05, то можно сделать окончательный вывод о существовании зависимости между этими двумя переменными.

Демонстрационный пример с использованием SPSS 12.2. Существуют ли различия в отношении к проблеме контроля над личным оружием между сторонниками разных политических идеологий?

Проблема контроля над оружием, находящимся в личном пользовании граждан, широко обсуждается в американском обществе. Можно ожидать, что либералы будут выступать в пользу контроля, консерваторы — противиться ему, а приверженцы умеренных политических взглядов займут некоторую промежуточную позицию. В файле с данными Общенационального социологического обследования — GSS 1998 поддержка контроля над личным оружием была выражена переменной *gunlaw (закон о контроле над оружием)*, а политическая идеология — переменной *polviews (политические взгляды)*. Последняя имеет слишком много категорий (7), чтобы быть [illegible] изменить разбиение этой переменной на категории. В главном меню выбрать пункт [illegible] Transform [illegible] вать в другую переменную). В данном примере преобразования переменная [illegible]

[illegible]

Согласно этой схеме [illegible] ческой позиции, все [illegible] переменной *pol*) [illegible] численное значение [illegible] ном между либ[illegible]

[illegible] Analyze, Descriptive Statistics [illegible]

Crosstabs Определить отношение [illegible] FAVOR OR OPPOSE GUN PERMITS ("за" или "против" разре- [illegible] Crosstabs [illegible] Percentages [illegible] Columns [illegible] симую переменную в строках [illegible] Statistics [illegible] ных *gunlaw* и *pol* выглядит следующим образом

Зависимость отношения к разрешению на хранение личного оружия от политических взглядов

		[illegible]			
		[illegible]	[illegible]	[illegible]	Всего
[illegible]	Процент от суммы по столбцу	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	Процент от суммы по столбцу	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Критерий хи-квадрат

	Значение	df (число степеней свободы)	Асимптотическая значимость (двусторонняя)
[illegible]	[illegible]	[illegible]	[illegible]
Отношение правдоподобия	6,780	2	[illegible]
[illegible]	[illegible]		

В каждой ячейке таблицы первым приведено количество получен[illegible]. Например, в первой [illegible] первого столбца указано число 235. Это количеств[illegible] которые поддерживают уста[illegible] [illegible] ного значения приведен проц[illegible] числа всех опрошенных лиц [illegible]

[illegible]

[illegible] случае, умеренная.

Демонстрационный пример с использованием SPSS 12.3. Направление связи. Изменяется ли отношение к сексу с возрастом?

Рассмотрим зависимость между двумя переменными, относящимися к порядковой шкале. Это упражнение позволит учащимся в дополнение к определению силы связи между двумя переменными приобрести опыт в описании направления такой связи. Обе необходимые для данного примера переменные должны принадлежать хотя бы к порядковой шкале. Рассмотрим в качестве независимой переменной *age (возраст)*. С помощью **Recode** (Преобразовать) разобьем массив значений этой переменной на три категории по аналогии с примером 10.1. Схема перехода к новой [illegible]

18 thru 34 → 1
35 thru 50 → 2
51 thru 89 → 3

[illegible] выборка разбита на три возрастных категории приблизительно одно- [illegible]

Исходя из обычного жизненного опыта, читатель может еще до анализа данных предположить, что связь между переменными *age (возраст)* и *premarsx (секс до бра[ка])* [illegible] [с]торонниками половых отношений до брака можно ожидать с его стороны [illegible] Crosstabs [illegible] [укажи]те строчную (зависимую) переменную как [illegible] *(секс до брака)*, и [illegible] независимую) переменную как *ager (воз[раст])* [illegible] (Это преобразо[illegible] прежняя переменная *age (возраст)*.) (Щелкн[ите]) [illegible] Cells (Ячейки) [illegible] [Отметьте], что в области **Percentages** (Проценты) [illegible] Columns (Столбцы) [illegible] по кнопке **Statistics** (Ста[тистики]) [illegible] [со]пряженности для переменных *premarsx* и *age*[r] [illegible]

Зависимость отношения к добрачным интимным отношениям от возраста респондента

		AGER			
		[illegible]	2	3	Всего
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

			AGE R			
			[illegible]	[illegible]	[illegible]	Итого
[illegible]	[illegible]	Количество	68	56	59	18[illegible]
	В этом нет [illegible]	Количество	149	166	80	395
Всего		Количество	267	310	274	871
		Процент от суммы по столбцу	100,0%	100,0%	100,0%	100,0%

Критерии хи-квадрат

	Значение	Степеней свободы	Асимптотический уровень значимости
Хи-квадрат Пирсона	54,350	6	[illegible]
Критерий линейной зависимости	54,239	1	0,000
Количество действительных случаев	871		

Нас интересуют следующие вопросы.

Существует ли связь между этими двумя переменными?

[illegible]

Если изучить таблицу от столбца к столбцу, то можно убедиться, что связь действительно существует. Положительна она или отрицательна? Вспомним, что в случае [illegible] высоким значениям частот другой, а низкие одной — низким второй. Отрицательная связь характеризуется противоположной зависимостью — высоким значениям частот одной переменной соответствуют низкие другой.

Найдем в каждом столбце ячейку с максимальным численным значением [illegible] столбец соответствует самой молодой возрастной группе. Для [illegible] самый [illegible]

[illegible]

ответом для некоторой переменной, приведены в приложении Ж или выводятся [illegible] Utilities — Variables [illegible]

[illegible] распространенным среди [illegible]

[illegible] рачному сексу [illegible]

расте. Значит, мы можем утверждать, что неприятие добрачных [illegible] увеличивается с возрастом [illegible] рассматриваемая связь [illegible]

[illegible] охарактеризовать как положительную, так и отрицательную в зависимости от точки зрения. Связь отрицательная, если рассматривать переменную *premarsx* как показатель одобрения, и положительная, если эту переменную воспринимать как показатель неприятия добрачного секса. Данный пример является иллюстрацией того факта, что в случае рассмотрения порядковых переменных, подобных *premarsx*, направление связи может быть неопределенным. Действительно, при положительной зависимости значения частот переменных всегда изменяются в одном направлении, а при отрицательной — всегда в противоположных. Однако по причине произвольности кодировки ординальных, или порядковых, переменных их высокие значения частот не обязательно означают возрастание измеряемого количества, а низкие численные значения не всегда означают уменьшение его количества. Поэтому всегда следует внимательно ознакомиться со таблицей кодировки переменных, а при анализе направления связи между порядковыми переменными проявлять осторожность.

Упражнения

12.1. Поскольку переменная *polviews (политические взгляды)* уже преобразована, вы можете на примере других общественных проблем (которым соответствуют такие переменные, как *cappun (смертная казнь)* или *busing*) проверить, согласуется ли характер связи с вашими предположениями (переменную *polviews* используйте как независимую). Действительно ли либералы занимают более [illegible] каждой таблицы следует запросить вывод процентных величин и значений критерия [illegible] наблюдаемую зависимость.

12.2. [illegible] отношение к ряду общественных проблем, [illegible] *cappun (смертная казнь)*, *grass (легализация марихуаны)* [illegible] *оружием)*, *busing*. Мы уже имеем результаты [illegible] в качестве независимой переменной *polviews* [illegible] поочередно принимая за независимую [illegible] *(пол)*, *relig (религия)* [illegible] *(общественный класс)* [illegible] (и с результатами [illegible] *polviews*) силу и направление [illegible] переменных. Сделайте [illegible] ной проблеме.

13 Связь между переменными, измеряемыми по номинальной шкале

Цели изучения

К концу этой главы вы сможете

1. Рассчитывать и интерпретировать коэффициенты фи (φ), V Крамера и лямбда (λ)
2. Объяснять суть подхода пропорционального уменьшения ошибки в терминах лямбды
3. Использовать любую из этих трех мер связи для анализа и описания зависимости между двумя переменными в контексте трех вопросов, которые рассматривались в главе 12

13.1. Введение

Меры связи являются описательными статистиками, которые обобщенно характеризуют силу связи между двумя переменными. Поскольку они представляют [illegible] в виде одного числа, то служат более эффективными показателями для выражения связи [illegible] связи, он упустит из виду определенную часть информации (детали и нюансы) [illegible] за правило наряду с обобщающей мерой связи всегда анализировать распределение частот [illegible] типа шкалы, к которой относятся измерения данных [illegible] меры связи.

[illegible]

13.2. Меры связи на основе критерия хи-квадрат

В течение многих лет специалисты по статистике в значительной степени полагались на меры связи, в основе которых лежит использование критерия хи-квадрат. Это связано с тем, что при известной величине критерия хи-квадрат вычисление мер связи не представляет труда. Для иллюстрации сказанного предлагается вернуться к рассмотрению таблицы 11.3, в которой на основе вымышленных данных показана связь между аккредитацией учебной программы и последующим трудоустройством для студентов, обучающихся по профилирующей дисциплине "социальные работники". Для удобства эта таблица воспроизведена в текущей главе под номером 13.1.

В главе 11 мы смогли убедиться, что данная взаимосвязь является статистически значимой ($\chi^2 = 10.78$ при уровне значимости $\alpha = 0.05$). Теперь нам следует выяснить силу связи. Даже из поверхностного рассмотрения таблицы 13.1 можно заключить, что условное распределение статуса трудоустройства действительно изменяется. Следовательно, статистическая связь между переменными существует. К этому выводу проще прийти, если рассчитать процентные величины для всех ячеек, после чего повысится наглядность представленных данных (см. таблицу 13.2).

Таблица 13.1. Зависимость трудоустройства выпускников профилирующей дисциплины "социальная работа" от аккредитации учебной программы (абсолютные величины; вымышленные данные)

	Статус аккредитации учебной программы		
Трудоустройство	Аккредитована	Не аккредитована	Всего
[illegible]	[illegible]	[illegible]	[illegible]
Трудоустроен как социальный работник	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Таблица 13.2. Зависимость трудоустройства от аккредитации (процентные величины; вымышленные данные)

	Статус аккредитации учебной программы	
Трудоустройство	Аккредитована	Не аккредитована
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

[illegible] переменными (аккредитация и трудоустройство) статистически значима [illegible] силы этой связи [illegible]

[illegible] коэффициент фи (φ). Эта [illegible] на основе критерия хи-квадрат [illegible] из двух строк и двух столбцов). [illegible] дующим образом:

Формула 13.1

[illegible] полученного результата.

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

[illegible]

Для таблицы 2 × 2 значение коэффициента фи может находиться в пределах от 0 (отсутствие связи) до 1,00 (полная связь). Величина $\phi = 0,33$ для таблицы 13.[illegible] говорит о существовании умеренной статистической связи между аккредитацией и трудоустройством. Значит, связь между двумя переменными статистически значима и является умеренной по силе. Что касается характера связи, то, как видно из табли[illegible]

[illegible] работ[illegible]

[illegible] мено и строк, и столбцов больше двух) верхний предел фи может быть выше 1,[illegible] [illegible] ающих размер 2 × 2, необходимо использовать более общую статистику, так назы[illegible] [illegible] Крамера. [illegible]

Формула 13.2

[illegible]

[illegible] минус 1) и [illegible] (число столбцов минус 1).

[illegible] следует найти квадратный корень из результата [illegible] на произведение N и меньшей величины из двух [illegible] минус 1. Максимальное значение [illegible]

[illegible]

С целью иллюстрации [illegible] Крамера [illegible] жим, что для выборки [illegible]

[illegible] ской успеваемост[illegible]

Таблица 13.3. Зависимость академической успеваемости студентов от их [illegible] в различных студенческих организациях (абсолютные величины)

Академическая успеваемость	Членство в организации			
	Братство (для мужчин) или женский клуб	Другая организация	Не состоит	Всего
Низкая	4	4	17	25
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Всего	23	26	26	75

Для этой таблицы хи-квадрат равен [illegible] с уровнем значимости 0,05. Находим коэффициент V Крамера.

[illegible]

Поскольку в таблице 13.3 одинаковое число столбцов и строк (по три), то в знаменателе будет произведение N на (3 − 1), т.е. 2. Рассчитаем [illegible] величина коэффициента [illegible]

[illegible]

[illegible] Члены мужских братств и женских клубов показывают среднюю успе[illegible] [illegible] нигде не состоит, — низкую.

Таблица 13.4. Зависимость академической успеваемости студентов от их членства в различных студенческих организациях (процентные величины)

Академическая успеваемость	Членство в организации		
	Братство (для мужчин) или женский клуб	Другая организация	Не состоит
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		
[illegible]			
	[illegible]		

[illegible] невозможность применения [illegible] 0,00 и 1,00. Обе эти меры являются показ[illegible] значениями, т.е. [illegible], например, величина 0,75 указывает на более сильную связь, [illegible]

[illegible]

[illegible] главы или главы 12 в следующих [illegible] рекомендуется рассмотреть задачи, приведенные в конце главы 11 [illegible]

[illegible]

Крамера имеют одинаковую величину.

[illegible]

В этой и в следующей главах наше основное внимание обращено на меры связи для двухвходовых таблиц сопряженности. Эти статистики очень широко используются для определения силы и — как будет показано в главе 14 — направления связи. Тем не менее первым шагом при анализе таких таблиц сопряженности всегда должно быть применение методов, описанных в главе 12. Процентные величины в столбцах и условные распределения могут сообщить больше подробностей о взаимосвязи между переменными, чем меры связи. Несмотря на их безусловную полезность, меры связи все-таки следует рассматривать только как обобщающие показатели, а не средства детального [illegible]

Но даже столь простые статистики, как проценты, могут быть рассчитаны с ошибками [illegible] лению позиции в профессиональных спортивных командах в зависимости от расы игрока [illegible] важность расчета процентных величин в правильном напра[illegible] Изучалась существующая практика назначения на отдельные позиции в [illegible] требуются лидерские качества или [illegible] [illegible] интерпретировались неверно, что [illegible] процентных величин.

[illegible] основанные на критерии χ^2, [illegible]

[illegible] что согласуется с сопровождающим ее текстом, — однако проценты ошибочно рассчитаны по строкам. Вторая таблица является примером правильного вычисления в направлении независимой переменной.

[illegible] циональном распределении игроков из различных этнических групп на определенных [illegible] торому разрешается занимать большинство позиций), — преимущественно работа для белых. Этот вывод (как мы знаем, ошибочный) может подтолкнуть к выдвижению какой-нибудь гипотезы, объясняющей отсутствие на этой не слишком престижной, но важной позиции в команде игроков из национальных меньшинств. Однако из рассмотрения правильной таблицы В следует, что инфилдерами в равной степени могут быть как белые, так и темнокожие и латиноамериканцы. Следовательно, данные таблицы А говорят скорее об этническом составе игроков высшей лиги, чем о том, какая расовая/этническая группа более полно представлена на позиции инфилдера.

Таблица А показывает процент игроков на каждой позиции в команде из каждой из рассмотренных расовых/этнических групп. Впечатление, что белые игроки доминируют на определенных позициях, создается лишь потому, что в таблице А они превосходят числом бейсболистов из других этнических групп.

В таблице В представлены процентные величины игроков в каждой группе, занимающие каждую из позиций в команде. Рассчитав проценты внутри столбцов, мы можем контролировать различия в размере групп и увидеть влияние расы (независимая переменная) на то, какую позицию (зависимая переменная) в команде занимает тот или иной игрок. Представители различных этнических групп распределены по большинству позиций в команде примерно поровну, но — что согласуется со структурой расового распределения — белые доминируют на позиции кэтчера (игрок, в первую очередь принимающий решения), а темнокожие в большинстве своем представлены на позициях аутфилдера (эти бейсболисты в наименьшей мере влияют на принятие решений).

Таблица А. Распределение в высшей бейсбольной лиге игроков различных рас по основным позициям в команде (процентные величины)[a]

Позиция в команде	Расовая принадлежность				
	Белые	Темнокожие	Латиноамериканцы	N	[illegible]
Кэтчер	87	[illegible]	[illegible]	[illegible]	[illegible]
1-я база	69	[illegible]	[illegible]	[illegible]	[illegible]
2-я база	59	[illegible]	[illegible]	[illegible]	[illegible]
3-я база	78	[illegible]	[illegible]	[illegible]	[illegible]
Шорт-стоп	42	[illegible]	[illegible]	[illegible]	[illegible]
Аутфилдер	34	48	[illegible]	[illegible]	[illegible]
Инфилдер	54	21	[illegible]	[illegible]	[illegible]

[a] *В отдельных случаях сумма значений не дает* [illegible]

Таблица 9. Распределение в высшей бейсбольной лиге игроков различных рас по основным позициям в команде (процентные величины)[1]

	Раса		
[illegible]	Белые	Темнокожие	Латиноамериканцы
[illegible]	7	5	4
1-я база	8	4	8
2-я база	7	3	4
[illegible]	8	2	17
[illegible]стоп	5	[illegible]	[illegible]
[illegible]			
Всего	99	101	10[illegible]
	(296)	(138)	(108)

В отдельных случаях сумма значений не дает 100% из-за погрешности округления. *Данные взяты из работы R. E. Lapchick and J. R. Benedict. 1993.* Racial Report Card. *Boston: Center for the Study of Sport in Society*

[illegible] *W. M. Leonard II and J. Phillips. 1997. "The Cause and Effect Rule for Percentaging Tables: An Overdue Statistical Correction for 'Stacking' Studies."* Sociology of Sport Journal, *14(5)*: 283 — 289

13.3. Пропорциональное уменьшение ошибки

В последние годы была разработана группа мер связи, основанных на подходе, известном как пропорциональное уменьшение ошибки (PRE). Они служат дополнением [illegible] PRE [illegible] самую переменную (X), попытаться угадать или предположить, в какую категорию зависимой переменной (Y) попадает каждое наблюдение. Поскольку предполо[illegible] [illegible] шого количества ошибок (то есть неверного предсказания категории многих [illegible]).

На втором этапе вновь делается прогноз категории каждого наблюдения [illegible] информация, которую [illegible] количество ошибок [illegible]

Поясним сказанное на примере. Предположим, вам поручили довольно необычное задание спрогнозировать, выше или ниже роста в 175 см будет каждый из случайных прохожих, встретившихся вам на улице, при условии, что вам совершенно ничего не известно об этих людях. Если эти встречные прохожие будут действительно случайными, то вы в ваших прогнозах будете ошибаться довольно часто (т.е. сказав, что следующий прохожий будет высоким, на самом деле повстречаете человека небольшого роста, и наоборот).

Теперь допустим, что вам предстоит пройти описанный тест дважды. Но на втором круге вам еще до того, как вы сделаете прогноз о росте, будут говорить, кем — мужчиной или женщиной — окажется очередной случайный прохожий. Поскольку рост связан с полом, человеку и женщины в среднем ниже мужчин, то оптимальной стратегией на втором круге прохождения теста может быть формулирование предположения, что все встреченные женщины ниже заданной отметки 175 см, а все мужчины — выше. Конечно, и в этом случае ошибок не избежать. Но, учитывая зависимость между двумя переменными (ростом и полом), можно ожидать, что во втором круге теста количество ошибок будет меньше, чем на первом круге. Другими словами, принимая во внимание информацию о независимой переменной, удается снизить количество ошибок прогнозирования. Для практического использования и сделанного наблюдения необходимо ввести некоторый статистический показатель [illegible]

13.4. Мера пропорционального уменьшения ошибки для номинальных переменных: лямбда

Результаты распределения 100 человек по росту и полу представлены в таблице 13.5. Даже без перехода к процентным величинам очевидно, что между переменными существует статистическая связь. Для измерения силы этой связи рассчитаем меру пропорционального уменьшения ошибки, которая называется лямбда и обозначается греческой буквой λ. Следуя подходу, о котором говорилось в предыдущем разделе, нам необходимо найти две величины. Первая — это количество ошибок про[illegible] прохожего), а вторая — также количество ошибок прогнозирования, но уже в слу[illegible]

Таблица 13.5. Зависимость роста встречного прохожего от его пола

	Пол		
Рост	Мужчины	Женщины	Всего
[illegible]	44	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Вначале проигнорируем инф[illegible] одному из полов и будем опер[illegible] В таком случае могут быть сдел[illegible]

[illegible]

[illegible] переменными связи при выполнении второй процедуры будет сделано ме[illegible] ошибок, чем при выполнении первой. Поскольку при учете половой при[illegible]сти субъектов нами было допущено меньше ошибок прогнозирования (14), че[м без] такового учета (48), то мы можно утверждать, что между рассматриваемыми двумя переменными (пол и рост) лонерника существует связь. Количество ошибок [illegible] снижено с 48 до 14. Пропорциональ[illegible]

Формула 13.3

[illegible]

[illegible] значение лямбды (0,71) [illegible] да, равная 0,00, означала бы, что информа[illegible] [не]зависимой переменной, не улучшила наши возможности по предсказа[illegible]

довательно, между двумя переменными отсутствует связь. С другой стороны, $\lambda =$ [illegible] [из]вестны значения X.

Популярность мер пропорционального уменьшения ошибки отчасти обусловле[на] [illegible] [интуитив]ный смысл. Эти значения можно рассматривать как показатель того, насколько X помогает нам предсказывать (или, если сформулировать более широко, понимать) Y. При умножении лямбды на 100 снижение количества ошибок и, следовательно, силы связи оказывается выраженным в процентах. Полученное в вышеприведенном примере значение лямбды следует понимать таким образом: знание о половой принадлежности случайного прохожего повышает наши возможности прогнозирования роста на 7[1] %. Или: наша уверенность при предсказании роста случайного прохожего будет на 7[1] % выше, если мы знаем половую принадлежность этого прохожего, чем в случае, когда мы не имеем такой информации.

13.5. Вычисление лямбды

Для обобщения процедуры вычисления коэффициента лямбда рассмотрим еще один пример. Допустим, что некоторый исследователь изучает зависимость между вероисповеданием человека и его отношением к смертной казни. В результате опроса 130 респондентов были собраны следующие данные.

Таблица 13.6. Зависимость отношения человека к смертной казни от его вероисповедания (вымышленные данные)

	Религиозная принадлежность				
Отношение к смертной казни	Католики	Протестанты	Другая	Неверующие	Всего
За	10	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Против	11	4	25	5	[illegible]
Всего	35	[illegible]	40	[illegible]	130

Шаг 1. Чтобы определить E_1, т.е. количество ошибок [illegible] [перемен]ной X (в данном случае, религиозной принадлежности) [illegible] [чи]сла наблюдений N максимальное итоговое значение [illegible]

$$E_1 = N - [illegible]$$

$$E_1 = [illegible]$$

$$E_1 = 80$$

Итак, если исследователь не [illegible] [принадл]ежность человека, то при оценке [illegible] [до]пустит ошибку в 80 случаях [illegible]

Шаг 2. Найдем E_2, т.е. кол[ичество] [illegible] [перемен]ной X. Для каждого [illegible]

Для католиков: 35 – [illegible] = [illegible]

Для протестантов: 25 – [illegible] = [illegible]

Для "других": 40 – 25 = 15;

Для неверующих: 30 – [illegible] = [illegible]

$E_2 = 65$

[illegible]

Шаг 3. На первом шаге было сделано 80 ошибок предсказания по сравнению с 65 на втором шаге. Поскольку количество ошибок уменьшилось, то можно уверенно утверждать, что между переменными существует связь. Для вычисления пропорционального уменьшения ошибки воспользуемся формулой [illegible]

[illegible]

[illegible] данном случае говорит о том, что с зависимой переменной связаны и другие, отличные от религии, факторы.

Практический пример 13.1

Случайную [illegible] разделить на две категории [illegible] их основной мотивацией к поступлению в [illegible] "профессиональная" или поиск хорошей работы и "академическая" [illegible] ситет было стремление [illegible] существует ли связь между этими двумя переменными [illegible]

Мотивация	Тип		
	Традиционные	Нетрадиционные	Всего
Профессиональная	25	60	[illegible]
Академическая	[illegible]	[illegible]	[illegible]
Всего	100	75	175

Для измерения силы связи между этими двумя переменными [illegible] ду

$$E_1 = 175 - 90 = 85;$$

$$E_2 = (100 - 75) + (75 - 60) = 25 + 15 = 40;$$

$$\lambda = \frac{E_1 - E_2}{E_1} =$$

$$\lambda = \frac{85 - 40}{85} = \frac{45}{85} = 0{,}53$$

Лямбда, равная 0,53, означает, что, прогнозируя мотивацию студентов с учетом их типа, исследователь сделал бы на 53% меньше ошибок, чем в случае прогнозирования без учета типа студента. Сила связи — от умеренной до значительной. Изучение таблицы позволяет сделать такой вывод: традиционные студенты в их стремлении к учебе с большей вероятностью будут руководствоваться академическими мотивами (75%), тогда как нетрадиционные больше ориентированы на повышение своего профессионализма и построение карьеры (80%).

13.6 Ограничения применения коэффициента лямбда

В качестве меры связи коэффициент лямбда имеет две характеристики, [illegible] следует обратить внимание. Во-первых, этот коэффициент является асимметричным [illegible] величина данной статистики разная в зависимости от того, [illegible] определить независимой и [illegible] связанных переменных [illegible] Например, в таблице 13.6 величина лямбды составила бы [illegible] независимой переменной [illegible] Следовательно [illegible] придерживаться правила помещения независимой переменной [illegible] и вычислять лямбду изложенным выше образом [illegible]

[illegible] коэффициента лямбда может [illegible] меры связи превышают 0,00 [illegible] получить $\lambda = 0{,}00$, даже когда другие меры связи показывают, что между переменными существует связь [illegible] для таблицы показывают [illegible] значительных перепадов между маргинальными [illegible]

[illegible] о на критерии хи-квадрат. *Для практичес[illegible] читатели могли рассмотреть любую из задач, при[illegible] главе 12 или 13. Как и в случае изучения коэффици[illegible] начинать с небольших выборок и таблиц 2 × 2.*

Заключение

1. [illegible] 13 были представлены три меры связи — коэффициенты фи, V Крамера [illegible] двумя переменными, помещенными в двухвходовую таблицу сопряженности.
2. Коэффициенты фи и V Крамера являются мерами, основанными на критерии хи-[illegible] коэффициент V Крамера может быть вычислен для таблиц любого размера. Хотя обе этих меры являются показателями силы связи, значения в пределах между 0,00 и 1,00 не имеют прямого и наглядного смысла.
3. Коэффициент лямбда является мерой, основанной на PRE — подходе, известном как **пропорциональное уменьшение ошибки**. Достоинством λ является наглядность интерпретации ее значений от 0,00 до 1,00. Величина статистики лямбда показывает, насколько точнее станет предсказание зависимой переменной, если имеются некоторые сведения о независимой переменной, по сравнению с точностью в случае, когда о независимой переменной ничего не известно. Благодаря наглядности интерпретации λ между двумя крайними значениями (т.е. между 0,00 и 1,00), этой мере связи часто отдают предпочтение перед более традиционными мерами, основанными на критерии хи-квадрат (за исключением ситуаций значительных перепадов между маргинальными частотами в строках таблиц).

Сводка формул

Фи (13.1) [illegible]

[illegible]

Словарь

Лямбда [illegible] PRE.

Пропорциональное уменьшение ошибки (PRE). [illegible] определение и вычисление коэффициента лямбда. Эта статистика является численным выражением сравнения количества ошибок прогнозирования за[illegible] количеством ошибок прогнозирования в случае принятия во внимание независимой переменной (E_2).

Фи (ϕ). Мера связи, основанная на критерии хи-квадрат. Применима для номинальных переменных, организованных в двухвходовую таблицу сопряженности.

E_1. Используется при вычислении лямбды и означает количество ошибок предсказания категории зависимой переменной, в которую попадают наблюдения в случае игнорирования независимой переменной.

E_2. Используется при вычислении лямбды и означает количество ошибок предсказания категории зависимой переменной, в которую попадают наблюдения в случае принятия во внимание независимой переменной.

V Крамера. Мера связи, основанная на критерии хи-квадрат. Применима к номинальным переменным, помещенным в таблицу сопряженности с любым числом строк и столбцов.

Мультимедийные ресурсы

The Wadsworth Sociology Resource Center: Virtual Society (Виртуальное сообщество Центра социологических ресурсов Вадсворта)

[illegible]

На web-сайте компании вы сможете найти шестое издание книги *Statistics: A Tool* [illegible] разнообразным ресурсам, полезным для тех, кто изучает [illegible] этой книги, вы найдете следующее:

- Обзор основ математики
- [illegible]
- Обучающие карты
- [illegible]
- Дополнительные задачи к отдельным главам
- Таблицу случайных чисел;
- Примеры и упражнения с использованием пакетов статистической обработки данных MicroCase и SPSS;
- Функциональные диаграммы "Find the test";
- Проверка гипотез для переменных, измеряемых по порядковой шкале.

Задачи

13.1 SOC [illegible]

[illegible] Затем несколько характеристик респондентов и полу- [illegible] дой из них требуется рассчитать фи и ламбда. Также следует дать ответ на вопрос: связь между какими переменными наиболее сильная?

а. Зависимость вероятности оказаться жертвой преступности от половой принадлежности респондента.

Вы подвергались ограблению?	Пол: Мужской	Женский	Всего
[illegible]	[illegible]	[illegible]	[illegible]
Нет	15	18	33
Всего	25	30	55

б. Зависимость вероятности оказаться жертвой преступности от возраста рес- [illegible]

Вы подвергались ограблению?	Возраст: До 21 года	[illegible] 22 лет	Всего
[illegible]	[illegible]	[illegible]	[illegible]
Нет	15	18	33
Всего	27	28	55

в. [illegible]

Вы подвергались ограблению?	Раса: [illegible]кожий	Белый	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	15	[illegible] 40	55

13.2. Для задач 12.1–12.4 расс[illegible]тать фи и ламбда. Сравнить полу- [illegible] рассмотрения процентных величин, рассчитанных в главе 12.

13.3 SOC

[illegible] самоубийства отчасти обусло[illegible] особенно среди молодежи. Возможно [illegible] ние, провоцируются мо- [illegible] средствах массовой информации. Был [illegible] ряд городов, которые разби[illegible] гории в зависимости от уров[illegible] чение года в местных СМИ м[illegible] ствует ли связь между [illegible] предложенных сфор[illegible]

Уровень самоубийств	Появлялись ли в СМИ материалы о самоубийствах? До	Нет	Всего
Низкий	15	20	35
Высокий	15	10	25
Всего	30	30	60

13.4 SW

Директор одного из приютов для избиваемых женщин заметила, что многие из ее подопечных в конце концов вновь возвращаются к своим жестоким мужьям, даже несмотря на очевидность того, что домашний тиран и впредь будет практиковать по отношению к своей жене рукоприкладство. Директор предположила, что причина покорности женщин очень проста: им просто больше некуда идти. Например, в данной местности у женщины нет ни одного близкого родственника, у которого жертва домашнего тирана могла бы временно пожить. Вопрос: подтверждают ли приведенные ниже данные предположение директрисы приюта? *Информация реальная, взята из личных файлов бывших обитательниц приюта.*

Вернулась ли женщина к мужу?	Есть поблизости родственники	Нет поблизости родственников	Всего
Да	[illegible]	33	[illegible]
Нет	50	7	57
Всего	60	40	100

13.5 PA

Традиционно в вашем городе автобусами пользовались люди с невысокими доходами и представители рабочего класса. Как председатель комитета городского общественного транспорта при мэрии, вы предположили, что жители при городов, в которых обитают преимущественно представители интеллигенции и среднего класса, будут чаще пользоваться автобусами, если маршруты последних соединят жилые пригороды с деловым центром города, куда многие добираются на работу. Было проведено социологическое исследование, результаты которого представлены в таблице. [illegible] связь [illegible] на автобусе с местоположением организации, в которой работает человек. [illegible] связь существует, то какова ее структура?

Потенциальный уровень пользования автобусом	Расположение места работы: Деловой центр города	Другое	Всего
Будет ездить на автобусе	55	20	[illegible]
Не будет ездить на автобусе	5	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

13.6. GER

Был проведен опрос пожилых людей [illegible] [illegible] для [illegible]

чении от условий их проживания?

Число правонарушений	[illegible]	Среди людей любых возрастов	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

13.7 SOC

Существует ли связь между половой принадлежностью преподавателя колледжа и той оценкой его/ее педагогических способностей, которую дают своему наставнику студенты? Требуется в нескольких предложениях сформулировать свои выводы.

	Пол преподавателя		
Педагогические способности	Женский	Мужской	Всего
Высокие	115	241	356
Низкие	54	[illegible]	167
Всего	169	354	523

13.8 SOC

[illegible] интимной близости до брака. Респондентам предлагалось выбрать один из двух [illegible] отражающие зависимость полученных ответов от нескольких переменных [illegible]

а. Зависимость отношения подростка к добрачной интимной близости от его [illegible]

	Пол подростка		
Добрачный секс	Мужской	Женский	Всего
Безусловно [illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	66	[illegible]

б. Зависимость отношения [illegible] к добрачной интимной близости [illegible] партнером

	Когда-нибудь встречались с „постоянным" партнером/девушкой?		
[illegible]	[illegible]	[illegible]	[illegible]
Допустим в определенных [illegible]	[illegible]	[illegible]	[illegible]

в. [illegible] ее принадлежности к общественному классу

	Общественный класс		
Добрачный секс	[illegible]	[illegible]	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

13.9 SOC

Один из колледжей США столкнулся с проблемой постепенной убыли числа студентов, которая выражается в том, что после очередных каникул некоторая часть учащихся не возвращается на занятия. Вопрос: существует ли связь между этим явлением и возрастом, расовой принадлежностью и формой обучения студентов? Для каждой из приведенных ниже таблиц требуется указать силу связи (если таковая существует) между убылью числа студентов и независимой переменной. Каков характер этой связи? *Подсказка: для ответа на данный вопрос рекомендуется рассчитать процентные величины.* Также требуется в виде нескольких предложений интерпретировать представленные в таблицах результаты опроса.

а. Зависимость убыли числа студентов от расовой принадлежности (исходная величина — 532 студента, посещавших занятия в осеннем семестре).

	Раса студента		
Убыль числа студентов	Белые	[illegible]	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

б. Зависимость убыли числа студентов от формы обучения (исходная величина — 532 студента, посещавших занятия в осеннем семестре).

	Форма обучения		
Убыль числа студентов	Заочная	Очная	Всего
Вернулся на весенний семестр	[illegible]	[illegible]	[illegible]
Не вернулся	[illegible]	[illegible]	[illegible]

в. [illegible] в осеннем семестре

	Возраст		
Убыль числа студентов	18–24	[illegible]	Всего
[illegible]	[illegible]	[illegible]	[illegible]
Не вернулся	[illegible]	[illegible]	[illegible]
Всего	[illegible]	[illegible]	[illegible]

[illegible] ...щее в обществе отношение к преступности. В [illegible]

[illegible]

Раса респондента

Уровень преступности	Белые	Чернокожие	Всего
[illegible]	[illegible]	[illegible]	[illegible]

13.11 [illegible]

[illegible] ...общественного класса рассматривают как основной. Помогут ли вам [illegible] ...решение о выборе средства массовой [illegible] ...для предвыборной агитации?

Основной источник новостей	Общественный класс		
	[illegible]	Белые воротнички	Всего
Радио	140	200	[illegible]
Газеты	25	40	[illegible]
Всего	[illegible]	[illegible]	[illegible]

13.12 [illegible] ...ти, взятые из отчета "Общенациональное социологическое исследование 1998 [illegible]

[illegible]

Вторая часть задания [illegible] ...но теперь независимой переменной является половая принадлежность рес[illegible]

[illegible]

а [illegible]

Пол

Право на аборт?	Мужской	Женский	Всего
[illegible]	[illegible]	[illegible]	[illegible]

б [illegible]

Пол

Смертная казнь?	Мужской	Женский	Всего
За	908	998	1906
[illegible]	[illegible]	[illegible]	[illegible]

в [illegible]

Пол

Право на эвтаназию?	Мужской	Женский	Всего
Да	[illegible]	[illegible]	[illegible]
Нет	[illegible]	[illegible]	[illegible]
Всего	[illegible]	[illegible]	[illegible]

г [illegible] ...ию (обозначение переменной — [illegible]

	Пол		
[illegible]	Мужской	Женский	Всего
[illegible]	[illegible]	[illegible]	[illegible]

[illegible] традиционного распределения ролей между [illegible] переменной – FEHOME) от пола респондента.

[illegible]	Пол		
[illegible]	Мужской	Женский	Всего
[illegible]	[illegible]	164	280
[illegible]	669	865	1534
[illegible]	785	1029	1814

Использование *SPSS for Windows* для вычисления мер связи между номинальными переменными

Демонстрационный пример с использованием SPSS 13.1. Существуют ли различия в отношении к проблеме контроля над личным оружием между сторонниками разных политических идеологий? Другой подход к нахождению ответа.

В разделе «Демонстрационный пример с использованием SPSS 12.2» для анализа [illegible] Crosstabs [illegible]

В данном примере мы вновь проанали[illegible] SPSS. [illegible] *Примечание: переменная* [illegible] Descriptive Statistics [illegible] чего переместить pol в поле [illegible] по Cells (Ячейки) и запрос[illegible] ста Percentages [illegible] Crosstabs [illegible] Columns

[illegible] и V Крамера [illegible] Щелкнуть Continue (Продолжи[ть]) и OK. При этом на экран будут выведены следующие таблицы.

Зависимость отношения к разрешению на хранение личного оружия от политических взглядов

			Политические взгляды			
			1.00	2.00	3.00	Всего
Зависимость отношения к [illegible] хранение личного [illegible]	"За"	Количество	235	247	239	7[illegible]
		Процент [illegible]				
	"Против"	Количество	35	50	64	149
Всего		Количество	270	297	303	870

Критерии хи-квадрат

	Значение	Степеней свободы	Асимптотический уровень значимости (двусторонний)
Критерий хи-квадрат Пирсона	6.724	2	0.035
Отношение правдоподобия	6.760	2	0.034
Критерий линейной зависимости	6.710	1	0.010
Количество действительных наблюдений	870		

Асимметричные меры связи

			Значение	Асимптотическая стандартная ошибка	Тау приблизительное	Уровень значимости приблизительный
Связь двух номинальных переменных	Лямбда	Симметричная [illegible] переменная "За" или "против" [illegible]	0.011	0.059	0.363	0.717

Симметричные меры связи

		Значение	Уровень значимости (приблизительный)
[illegible]	Фи	0,088	[illegible]
	[illegible]	0,088	[illegible]
Количество действительных наблюдений		870	

[illegible]

[illegible] FAVOR GUN PERMITS [illegible]

[illegible] Другими словами, политическая [illegible] человека не играет особой роли в его отношении к проблеме контроля [illegible] оружием.

Демонстрационный пример с использованием SPSS 13.2. От каких переменных зависит поддержка контроля над личным оружием?

[illegible] другие из обычно используемых [illegible]

[illegible]

[illegible] Использование [illegible] переменных (хотя также [illegible]) [illegible] обрабатывать большие объемы информации). В данном примере для [illegible] будет использован сокращенный вариант [illegible] при помощи системы SPSS

	% выступающих "за"*	Уровень значимости критерия хи-квадрат	Фи* или V	Лямбда
ПОЛ				
Мужской	[illegible]			
Женский	88,5	0,000	0,15	0,00
РЕЛИГИЯ				
Протестант	[illegible]			
Католик	[illegible]			
Иудей	[illegible]			
Неверующий	[illegible]			
Другая	[illegible]	[illegible]	[illegible]	[illegible]
РАСА				
Белый	81,6			
[illegible]	[illegible]			
Другая	[illegible]	[illegible]	[illegible]	[illegible]
СЕМЕЙНОЕ ПОЛОЖЕНИЕ				
Женат/Замужем	83,6			
Вдовец/Вдова	89,1			
В разводе	82,7			
Временно живут отдельно	82,4			
Не был женат/Не была замужем	81,4	0,514	0,06	0,00

* В некоторых случаях фи выводится на экран со знаком "–". При анализе таблицы знак можно [illegible]

В [illegible]

[illegible] значений коэффициента лямбда.

Упражнения

13.1 [illegible] SPSS Recode [illegible] MicroCase Collapse Variable [illegible]

13.2 [illegible]

14 Связь между переменными, измеряемыми по порядковой шкале

Цели изучения К концу этой главы вы сможете

1. Рассчитывать и интерпретировать коэффициенты гамм [illegible]
2. Объяснять логику пропорционального уменьшения о[illegible] на основе коэффициента гамма
3. [illegible] лиза и описания зав[illegible] контексте тех трех вопросов, о которых говорилось в [illegible]
4. Проверять значимость коэффициентов гамма и ро Спирм[illegible]

14.1. Введение

[illegible] пользования шкалы отношений, которая [illegible]

Порядковые переменны[illegible] [illegible] енных значений (не более [illegible] набора данных в свернутой форме [illegible] алы. Например, [illegible] сколькими категорий (таких как высокая, средняя, низкая).

Для работы со свернутыми поря[illegible] [illegible] ми Спирмена [illegible] данными (такие как [illegible]

[illegible] ро Спирмена [illegible]

Данная глава расширит понимание читателями способ[illegible] [illegible] Существует ли между переменными связь? Насколько сильна эта связь? Каково направление связи?

14.2 Пропорциональное уменьшение ошибки (PRE)

В случае номинальных переменных подход пропорционального уменьшения ошибки базировался на двух этапах предсказания. На первом этапе мы предсказывали Y, игнорируя X, а на втором этапе мы предсказывали Y уже с учетом X. Количественной мерой пропорционального уменьшения ошибок предсказания на втором этапе по сравнению с первым является коэффициент лямбда.

Для порядковых переменных также используются два аналогичных этапа предсказания, а мерой пропорционального уменьшения ошибки выступает коэффициент [illegible]ма. Основное различие между подходом пропорционального уменьшения ошибки для номинальных и порядковых переменных заключается в том, каким об[illegible] предсказания. В случае гаммы прогнозируется порядок пары наблюдений. Другими словами, мы прогнозируем, будет ли значение одного наблюдения изучаемой переменной находится выше или ниже значения второго наблюдения. На первом этапе выполняется предсказание взаимного расположения пары наблюдений одной переменной, игнорируя их порядок для другой переменной. На втором этапе предсказывается взаимного расположения для одной переменной с одновременным [illegible] порядка для другой переменной.

Для примера предположим, что исследователя заинтересовали случаи "перегора[illegible] [illegible] ей начальных школ (т.е. случаи, когда учитель теряет интерес к своей работе). Требуется проверить гипотезу, что "перегорание" учителя зависит от стажа его/ее работы по этой специальности. Сформулируем вопрос по-другому: сильнее ли выражено "перегорание" у тех учителей начальной школы, у которых более про[illegible]тельный стаж работы по специальности? Если нам известно, что учитель А проработал до[illegible] учитель В, то можем ли мы точно предсказать, что и степень "перегорания" учителя А выше, чем учителя В? То есть, поможет ли нам зна[illegible] о взаимном расположении пары наблюдений одной переменной предсказать взаимное расположение пары наблюдений для другой переменной?

[illegible]сли между двумя переменными существует связь, то мы сможем уменьшить [illegible] связи между переменными гамма равна 0,00 [illegible] наблюдений одной переменной не улуч[illegible] положении и пары наблюдений дру[illegible]

[illegible]

[illegible] стажем полож [illegible] При отрицательной связи порядок следова [illegible]

[illegible]

[illegible] второй переменной оно будет расположено ниже наблюдения В. Небольш [illegible]

[illegible]

14 3. Вычисление коэффициента гамма

В таблице 14.1 приведены данные о связи между стажем работы и степенью [illegible] регорания" для вымышленной выборки из [illegible] учителей начальной [illegible] числения гамма вначале требуется определить две суммы. Во-первых, след [illegible] ти число пар наблюдений согласованно упорядоченных для обеих переменных [illegible] величину обозначим N_s). Затем находим число пар наблюдений рассогласованно упорядоченных для этих же переменных (N_d). Обе величины можно определить, [illegible] из приведенных в ячейках таблицы частот наблюдений

Таблица 14.1. Зависимость степени "перегорания" учителей начальной школы от продолжительности трудового стажа по специальности (абсолютные величины, вымышленные данные)

Степень "перегорания"	Трудовой стаж			Итого
	Короткий	Средний	Большой	
[illegible]	[illegible]			[illegible]
[illegible]		6	3	[illegible]
[illegible]		5	5	[illegible]
[illegible]		[illegible]	21	[illegible]
Итого	38	[illegible]	30	[illegible]

[illegible] наблюдения [illegible] такая ячейка расположена в верхнем левом [illegible] ранг наблюдения слева направо по строке и [illegible] следует убедиться [illegible] указанные в верхней [illegible] низкой степени "перегорания" [illegible] для обозначения ранга любой пер[illegible] вычислении будем ис[illegible]

[illegible]

Прежде это [illegible] обозначим наблюдения в верхней левой ячейке, как НН ("низ-[illegible]

Далее формируем пару наблюдений [illegible]

[illegible]

Любая пара наблюдений, сформированная между этими двумя ячейками (верх[illegible] левой и центральной), согласованно упорядочена по обеим переменным. То есть, все Н[illegible] ниже всех СС по обеим переменным (как по шкале X, так и по шкале Y ранг [illegible] ранг действительно ниже "среднего" ранга). Число таких возможных пар [illegible] определяется произведением частот этих двух ячеек. Так [illegible] двух ячеек общее число N [illegible] = 300.

Гамма игнорирует все пары наблюдений, имеющие связи* по любой из двух переменных. Например, любая пара наблюдений, сформированная между НН и любой другой ячейкой верхней строки (низкая степень "перегорания") или левого стол[illegible] (низкий трудовой стаж), будет иметь связку по одной переменной. К тому же [illegible] пара наблюдений, сформированная на основе наблюдений одной ячейки, будет иметь связку как по X, так и по Y. Коэффициент гамма игнорирует все пары на[illegible]блюдений, сформированных на основании наблюдений из одной и той же строки, столбца или ячейки. На практике это означает, что при вычислении N_s мы будем учитывать только те пары наблюдений, которые могут быть сформированы между каждой ячейкой и ячейками, расположенными ниже и правее ее.

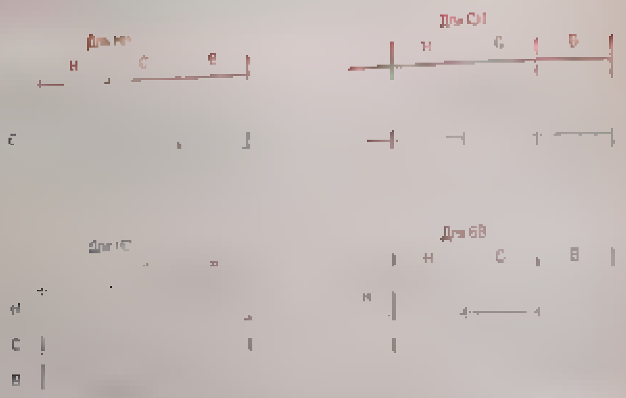
Вывод: для нахождения общего числа всевозможных пар наблюдений, и согласованно упорядоченных по обеим переменным (N_s), следует умножить частоту каждой ячейки на сумму всех частот в ячейках, расположенных от нее ниже и справа. Повторить описанную процедуру для всех ячеек, затем сложить полученные произ[illegible] (сумма представляет собой N_s). Выполнение процедуры для каждой ячейки [illegible] 14.1 приведено ниже. Следует заметить, что ячейки нижней строки и правого столбца не вносят вклада в N_s, поскольку ниже и справа от них ячейки отсутствуют. На рис. 14.[illegible] изображено направление умножения для каждой из четырех [illegible], которые могут (в случае таблицы 3×3) сделать вклад в N_s. Вычисляя N_s для таблицы 14.1, мы обнаружим, что согласованно упорядочены по обеим переменным [illegible] общей сложности [illegible] пара

[illegible] N_s

[illegible]	[illegible]		
Для СН	6×(21 + 5)	=	156
[illegible]			
Для НС	[illegible]×(5 + 21)	=	320
Для СС	(5×21)	=	105
Для ВН	3×(0)	=	0
[illegible]	[illegible]	=	0
[illegible]			

[illegible]

* [illegible] значений переменной, ранги которых одинаковы. — *Прим. науч. ред.*



[illegible]

$$N_r = 499$$

Формула 14.1

$$G = \frac{N_s - N_d}{N_s + N_d}$$

[illegible] переменным [illegible]

[illegible] переменных [illegible]

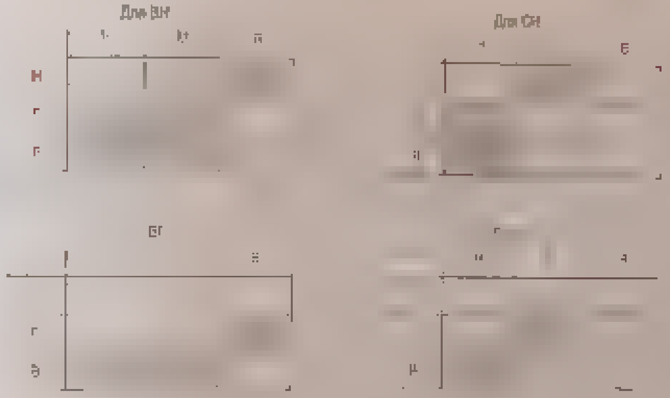


Рис. 14.2. Вычисление N_r для таблицы 3x3

Вычислим гамму для таблицы 14

$$G = \frac{1831 - 499}{1831 + 499}$$

$$G = \frac{1332}{[illegible]}$$

$$G = 0,57$$

[illegible]

...дать вывод о существовании связи между [illegible] ...теля и степенью его "перегорания", а о том, что [illegible]

[illegible] наблюдения А также выше, чем степень перего[illegible]

Для использования предложенного выше порядка вычисления гаммы требуется [illegible] обную таблице 14 [illegible]. То есть необходимо, чтобы [illegible]

[illegible] обходимо только по такой схеме [illegible]

[illegible] направление. Гамма — симметричная мера связи, т.е. ее значение не зависит от того, какую из двух переменных принять за независимую. *Для практического освоения [illegible] по 14.10 и 14.15. Начинать следует с малых таблиц (2×2), поскольку их использование упростит процесс освоения процедуры вычисления.*

Практический пример 14.1

Повышается ли активность избирателей на местных выборах, если кандидаты расходуют больше денежных средств на политическую рекламу в средствах массовой информации? Каждому из 177 рассмотренных избирательных участков был присвоен ранг в зависимости от явки избирателей ("высокая" или "низкая"), а полные избирательные бюджеты кандидатов ранжированы как "высокие", "средние" и "низкие". Вопрос: существует ли связь между этими двумя переменными?

Избирательный бюджет кандидата

Явка избирателей	Низкий	Средний	Высокий	[illegible]
Низкая	35	32	17	[illegible]
Высокая	23	27	43	93
Итого	58	59	60	177

Поскольку обе переменные — порядковые, то для оценки тесноты и направления связи удобно вычислить коэффициент гамма. Число пар наблюдений, согласованно упорядоченных в одинаковом порядке по обеим переменным [illegible]

$$N_s = 35 \times (27 + 43) + 32 \times (43) = 2450 + 1376 = 3826$$

Число пар наблюдений, рассогласованно упорядоченных по двум переменным [illegible]

$$[illegible]7 \times (27 + 23) + 32 \times (23) = [illegible]$$

Вычисляем гамму

$$G = [illegible]$$

$$[illegible] \quad 3826 + 1586$$

$$G = \frac{2240}{5412}$$

$$G = [illegible]$$

[illegible] ществования между двумя переменными средней по силе и положительной свя[illegible]

Соответствующие статистики

Статистики, связанные с любой двумерной таблицей сопряженности, обычно приводят непосредственно под таблицей. Эта информация содержит название меры связи, ее значение и знак (если данная мера может принимать положительные и отрицательные значения). Кроме того, если исследованием охвачена случайная выборка, то приводятся еще величина критерия хи-квадрат или уровня значимости самой меры (см. раздел 14.7). Следовательно, информация, приводимая под таблицей, может выглядеть таким образом

$$[illegible]$$

$$\chi^2 = 13{,}23\ (2\ df,\ p < 0{,}05)$$

Уровень значимости (альфа) приводится в виде "$p <$", который обсуждался в разделе "Чтение статистик 5".

Как правило, авторы статьи не просто приводят значения указанных статистик, но также дают в тексте их словесную интерпретацию. Так, вышеприведенные статистики могут быть интерпретированы следующим образом: "Связь между переменными статистически значима и умеренна по силе". Обычно исследователи избегают многословного стиля описания результатов (несмотря на большую наглядность и точность этого приема). Там, где читатели данной книги уже вооруженные почерпнутыми из нее знаниями могли бы сказать: "... Лямбда, равная 0,47, говорит о том, что по сравнению с предсказанием зависимой переменной без учета независимой мы уменьшили бы количество ошибок предсказания на 47%, если бы давали прогноз зависимой переменной с учетом информации о независимой переменной", исследователь ограничился бы таким [illegible] что его аудитория достаточно разбирается в статистике и потому может самостоятельно сформулировать более детальные интерпретации.

Профессиональная литература по статистике

В последние десятилетия Соединенные Штаты [illegible]

[illegible]

...тагов и приведены характеристики связи между несколькими демографическими и дву-
...зависимыми переменными. Первой зависимой переменной измерялась [illegible]

[illegible]

...ких вопросов очень ... Менее обеспеченные ...

Таблица 1. Влияние демографических переменных ... отношения респондентов к этическому вопросу

Белые	2.09	[illegible]	[illegible]	[illegible]
Темнокожий	1.96	[illegible]	[illegible]	[illegible]
Доход	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

14.4. Определение направления связи

[illegible] ...минус указывает на отрицательную связь. В то же время при работе с порядковыми [illegible] ...поэтому мы рассмотрим данный вопрос детально. Вначале остановимся на [illegible]

[illegible]

Таблица 14.2. Обобщенная форма положительной связи

	Переменная X		
Переменная Y	Низкий	Средний	Высокий
Низкий	X		
Средний		X	
Высокий			X

В таблице 14.3 представлена положительная связь с использованием реальных данных. Выборка составлена из 186 стран со всего мира доиндустриальной эпохи. Каждой стране была присвоена оценка, отражающая степень социального расслоения, и дано определение политического устройства, принятого в этой стране. Страны с высокой степенью социального расслоения характеризуются очень большими [illegible] между представителями высших и низших слоев общества по уровню [illegible] и обладания властью. С другой стороны, в странах, где наблюдает[illegible] расслоение, практически отсутствуют различия между [illegible] состояния и обладания властью. Следует отметить направ[illegible]ной: неравенство возрастает слева направо.

Таблица 14.3. Зависимость политического устройства страны от степени социального [illegible]

	Степень социального расслоения		
Тип политического устройства страны	Низкая	Средняя	Высокая
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible] образом [illegible] принимаются согласованным мнением всех [illegible]. Страны, в которых существует государственное ус[illegible]ительство. Наконец, страны с "отдельными эле[illegible] занимают промежуточное положение [illegible]

[illegible]

[illegible] переменными.

Отрицательные взаимосвязи являются противоположностью положительных [illegible] значения одной пер[illegible] уменьшаются вдоль диагонали от нижней левой до верхней правой ячейки (во всяком случае, это справедливо для всех таблиц данной книги). В таблице 14.4 пред[illegible] обобщенная форма отрицательной связи [illegible] значениями по переменной X, как правило, принимают более низкие значения по переменной Y, значения Y уменьшаются при возрастании значений X.

Таблица 14.4. Обобщенная форма отрицательной связи

	Переменная X		
Переменная Y	Низкий	Средний	Высокий
Низкий			X
[illegible]		[illegible]	
[illegible]	[illegible]		

Таблица 14.5 является примером отрицательной связи с использованием [illegible] данных опроса репрезентативной выборки граждан США, взятых из отчета [illegible] национальное социологическое исследование 1998 года[*]. В качестве [illegible] переменной принята посещаемость церкви, а зависимая переменная отражает [illegible] степень одобрения респондентом сожительства вне брака (т.е. ответ на вопрос [illegible] ли вы совместное проживание мужчины и женщины [illegible] заключения брака?"). *Примечание: частота посещения [illegible] направо; степень одобрения сожительства вне брака [illegible]*

Таблица 14.5. Зависимость степени одобрения сожительства вне брака[a] от частоты посещений церкви

	Частота посещений церкви		
Степень одобрения сожительства вне брака	Никогда	Раз в неделю или реже [illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

Как и в предыдущем примере, характер связи в таблице очевиден даже без пере- [illegible] Значительное большинство людей [illegible] редко посещаю- [illegible]

[illegible] добросовестных прихожан не были склонны одобрять такой вы- [illegible]

[illegible] может означать как "больше", так и "меньше". Например, ранжи- [illegible]

[illegible] номера этим категориям любым из двух способов

A	B
[illegible]	[illegible]
(2) Средний	(2) Средний
[illegible]	[illegible]

Несмотря на то, что схема B выглядит предпочтительнее, поскольку более высо- [illegible] значениям переменной соответствует и более высокий номер категории, [illegible] схемы абсолютно правильны и законны, однако направление связи будет мен- [illegible] в зависимости от выбора схемы (A или B). При использовании схемы B [illegible]

[illegible]

14.5. Интерпретация статистической связи с помощью таблиц сопряженности на примере исследования участия граждан США в общественной жизни

До настоящего момента [illegible] помощи таблиц, столбцы [illegible] [illegible] примере исследования участия гр- [illegible]

[illegible] американцев из общественной [illegible]

[illegible] общественной жизни, постоянно сокра- [illegible] Это сложная и многогранная проблема, и мы ее [illegible] в нескольких параграфах [illegible] данные Общего [illegible] 1998 года, мы сможем изучить источники [illegible]

Для измерения участия в общественной жизни [illegible] можно [illegible]

[illegible] по-разному. Несмотря на [illegible] абсолютно точно измерить участие человека в общественной жизни. [illegible]

[illegible]

[illegible] жизни и каждым из трех факторов, и каков характер или направление этих свя- [illegible]

[illegible] Зависимая переменная — [illegible]

[illegible] примере будем [illegible]

[illegible] только номинальной [illegible] категории [illegible]

[illegible]

...коэффициент гамма, а для номинальных — лямбда или V Крамера. Однако такое разделение не создает проблему сравнимости: сможем ли мы сказать, какая связь сильнее, если будем использовать разные меры? Есть два пути решения этой проблемы. Во-первых, для сравнения между собой разных таблиц можно воспользоваться столб... [illegible] Во-вторых, [illegible] ...ные меры связи.

[illegible]

...утверждение, будто женатые люди являются источником оздоровления общественной [illegible]

Таблица [illegible]. Зависимость участия в общественной жизни от семейного положения

Вы являетесь членом какой-нибудь общественной организации?	Семейное положение		
	[illegible]	[illegible]	[illegible]
Нет	37,7%	[illegible]	[illegible]
Да	62,3%	[illegible]	[illegible]
Итого	100%	[illegible]	[illegible]

[illegible]

Зависимость степени участия в общественной жизни [illegible] (таблица 14.7) не вполне статистически значима [illegible]

[illegible]

...Полученные данные результаты не позволяют [illegible] ...причинной связи [illegible]

[illegible] ...сопряженная связь — разные понятия.

Таблица 14.7. Зависимость участия в общественной жизни от занятости на работе

Вы являетесь членом какой-нибудь общественной организации?	Занятость на работе	
	Полная	Частичная
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

Таблица 14.8. Зависимость участия в общественной жизни от количества времени просмотра телепередач

Вы являетесь членом какой-нибудь общественной организации	Длительность телепросмотра в день		
	Низкая (0–1 час)	Средняя (2 часа)	Высокая (более 3 часов)
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Можно сделать следующий вывод: люди, состоящие в браке, и те, кто работает [illegible]

[illegible]

14.6. Коэффициент ро Спирмена (r_s)

[illegible] ...в таблицах. Однако, для многих порядковых переменных характерен широкий [illegible] ...ция. Такие данные можно свернуть до нескольких более широких категорий (напри-

при помощи коэффициента гамма или других мер связи для свернутых

...

Далее исследователь захотел бы проверить утверждение, что

...

таблице представлены в таблице 14.5

Таблица 14.5

...

Далее можно свернуть данные и построить таблицу сопряженности. К примеру, ... на две части и таким образом ... значения: высокое и низкое. Хотя подобное сворачивание ... допустимо, а часто даже необходимо ...

...

Ро Спирмена (r_s) является мерой ...

любой из переменных. Математическая обработка ... ных, конечно, невозможна, а только лишь допустимы утверждения вида "больше чем" и "меньше чем". Поэтому для вычисления показателя ро Спирмена наблюдения значение располагаются в убывающем порядке по обеим переменным от больших значений к меньшим ... В таблице 14.16 приведены для обоих переменных исходные значения и ранжи наблюдений.

Дебби	7	2	18	1	1	1
Эвелин	10	5	8	8		
Зигфрид	1	1				

мальное значение по X, поэтому ей присваиваем ранг 1 по этой переменной. Затем все остальные наблюдения ранжируются в убывающем порядке. Наблюдениям, ко...

Присваиваем ранг, равный среднему арифметическому тех рангов, которые получили бы эти переменные, если бы их значения не совпадали. К примеру, для Кристи и ... бы эти два наблюдения по переменной X не совпадали, то им были бы присвоены ...

... такой ранг присваивается обоим совпадающим наблюдениям по переменной X. (Аналогично рассчитывается ранг для более чем двух совпадающих наблюдений ...)

... гов (7, 8 и 9). Ро Спирмена вычисляется по следующей формуле

Формула 14.2

где D — квадрат суммы разно...

Для определения $\sum D^2$ сначала необходимо для каждого наблюдения вычесть ранг [illegible] D является разностью между рангами по X и Y). В таблице 14.10 для D предусмотрен отдельный столбец. Следует заметить, что сумма по этому столбцу ($\sum D$) равна 0. Значит сумма отрицательных разностей между рангами равна сумме положительных разностей. Поскольку такой результат получается всегда, то суммирование по столбцу D может служить проверкой правильности вычислений до данного момента. Если выяснится, что $\sum D$ не равна 0, значит, вами была допущена ошибка либо при ранжировании наблюдений, либо при нахождении разностей между рангами по двум переменным.

В столбце D^2 каждая разность D возводится в квадрат для устранения отрицательных знаков. Сумма по столбцу ($\sum D^2$) подставляется непосредственно в формулу 14.2:

$$r_s = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

[illegible]

$$r_s = 1 - \frac{[illegible]35}{990}$$

[illegible]

Ро Спирмена является показателем тесноты связи между переменными [illegible] этот к[illegible]

[illegible] ющая отрицательная связь ($r_s = -1{,}00$) существует при полном несоответствии между рангами по двум пере[illegible] переменной и т.д.). Ро Спирмена [illegible]

[illegible] ения ошибки. Показатель ро-квадрат (r_s^2) отража[illegible] ошибки предсказания ранга одной переменной [illegible] по сравнению с пр[illegible] В рассмотренном выше при[illegible]

$r_s = 0{,}86$, поэтому $r_s^2 = 0{,}74$. Так, [illegible] моуважения была бы меньше на 74%, если бы мы [illegible]

[illegible] приложенные с ней

Практический пример 14.2

Выборка из пяти крупных городов была ранжирована по индексу качества жизни. Также был определен процент населения, переехавшего в город для постоянного проживания за последний год. Вот вопрос: привлекают ли больше новых жи[illegible] сведены в таблицу. Каждому наблюдению присвоен ранг по обеим переменным, рассчитаны разности рангов и квадраты разностей.

Город	Качество жизни	Ранг	Процент новых жителей	Ранг	D	D^2
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
E	2	5	5	4	[illegible]	[illegible]
					0	4

[illegible]

$$r_s = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

[illegible]

[illegible] качество жизни, тем выше процент новых жи[illegible] [illegible] ой переменной с учетом ранг[illegible] [illegible] рангов другой переменной.

14.7. Проверка нулевой гипотезы об отсутствии связи посредством коэффициентов гамма и ро Спирмена

Работая со случайными выборками и с сформированными [illegible] полученные результаты на генеральную совокупность. Во [illegible]

например, разность между двумя выборочными средними — для генеральных совокупностей, из которых были извлечены эти выборки. Проверка [illegible] зависимо от ее формы и используемого критерия [illegible] на вопрос, существуют ли в генеральной совокупности [illegible] или разности, или зависимости, которые [illegible] значимость могут быть подвергнуты и меры связи [illegible] либо случайной выборке, то нам требуется не только [illegible] связи и измерить ее силу [illegible] ление о существовании связи между переменными [illegible]

[illegible]

но провести для таблиц, отражающих зависимость между [illegible] порядковыми переменными. Однако критерий хи-квадрат [illegible] показывает, что именно [illegible]

[illegible] может использоваться [illegible] проверки значимости собственно меры связи (гаммы или ро Спирмена).

При проверке гаммы или ро Спирмена на статистическую значимость нулевая гипотеза утверждает, что в генеральной совокупности отсутствует связь между [illegible] меры связи для генеральной [illegible] Значения мер связи для генеральной совокупности [illegible] ваны гамма (γ) и ро (ρ). Для обеих мер связи проверка [illegible] уже знакомого нам пятишагового алгоритма [illegible]

Демонстрация проверки значимости [illegible] гаммы [illegible] примере [illegible] 14.1, для которой гамма равна 0,57.

Шаг 1. Принятие предположений

Когда размер выборки [illegible]

[illegible]

Модель: Случайная выборка
Переменная измерена по порядковой шкале
Нормальное распределение выборочных статистик

Шаг 2. Формулирование нулевой гипотезы

[illegible]

Шаг 3. Выбор распределения выборочных статистик и задание критической области

Для выборок из 10 наблюдений и более Z-распределение (приложение А) может использоваться для поиска областей под распределением выборочных статистик.

Распределение выборочных статистик = Z-распределение

[illegible]

Шаг 4. Вычисление статистики критерия

[illegible]

$$Z\text{ (экспериментальное)} = 0{,}57\sqrt{\frac{[illegible]+499}{[illegible](1-0{,}33)}}$$

$$Z\text{ (экспериментальное)} = 0{,}57\sqrt{\frac{2330}{[illegible](0{,}67)}}$$

$$Z\text{ (экспериментальное)} = 0{,}57\sqrt{34{,}78}$$

$$Z\text{ (экспериментальное)} = (0{,}57)(5{,}90)$$

[illegible]

Шаг 5. Принятие решения

Сравним Z (экспериментальное) и Z (критическое):

$$Z\text{ (экспериментальное)} = 3{,}36$$

$$Z\text{ (критическое)} = \pm 1{,}96$$

[illegible] ваемые переменные связаны между собой также для генеральной совокупности [illegible]

[illegible] задачи 14.2, 14.4, 14.7, 14.10 и 14.15.

Проверку на основе показателя ро Спирмена можно выпол[illegible] ральной совокупности (ρ) фактически равно [illegible]

Спирмена приближается к [illegible] зовать именно этот вид распределения [illegible] Спирмена выполнена на примере раздела 14.6.

Шаг 1. Принятие предположений

Модель: Случайная выборка.
Переменная измеряется по порядковой шкале.
Нормальное выборочное распределение.

Шаг 2. Формулирование нулевой гипотезы

[illegible]

Шаг 3. Выбор распределения выборочных статистик и задание критической области

Распределение выборочных статистик = [illegible]
Альфа = [illegible]
Число степеней свободы = [illegible]
t (критическое) = [illegible]

Шаг 4. Вычисление статистики критерия

[illegible]

t (экспериментальное) = 3,86

[illegible]

Шаг 5. Принятие решения

[illegible]

t (экспериментальное) = 4,77

Видно, что нулевую гипотезу [illegible] отвергнуть. С вероятностью ошибки [illegible] можем сделать вывод, что рассматриваемые переменные [illegible] же во всей генеральной совокупности [illegible] которой была извлечена выборка. [illegible] выполнения и интерпретации проверки на значимость [illegible] Спирмена читателям рекомендуется решить задачи с 14.11 [illegible]

Заключение

1. В этой главе были рассмотрены такие меры связи, как коэффициент гамма для свернутых порядковых переменных и коэффициент ро Спирмена, который применяется для "непрерывных" порядковых переменных. Обе меры указывают на силу и направление связи между переменными.
2. Гамма — это мера связи, основанная на подходе пропорционального уменьшения ошибки. Она показывает улучшение нашей способности предсказывать порядок пар наблюдений по одной переменной с учетом порядка пар наблюдений по другой переменной по сравнению с предсказанием, когда порядок пар наблюдений по другой переменной игнорируется.
3. Ро Спирмена вычисляется на основе рангов наблюдений по двум "непрерывным" порядковым переменным и при возведении во вторую степень может интерпретироваться при помощи подхода пропорционального уменьшения ошибки.
4. В случае вычисления гаммы и ро Спирмена для случайной выборки, извлеченной из определенной генеральной совокупности, обе меры связи должны быть проверены на статистическую значимость. Нулевая гипотеза — это утверждение об отсутствии связи между двумя переменными в генеральной совокупности. Проверка на статистическую значимость выполняется следуя уже известному пятишаговому алгоритму.

Сводка формул

Гамма	14.1	$G = \frac{N_s - N_d}{N_s + N_d}$
Ро Спирмена	14.2	$r_s = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$

Словарь

Гамма (G). Мера связи, которая может применяться для свернутых порядковых переменных, представленных в табличном формате. Символом G обозначается выборочная гамма, а символом γ — гамма для генеральной совокупности.

Ро Спирмена (r_s). [illegible] применяется для "непрерывных" [illegible] символом ρ_s [illegible]

N_s. Число пар наблюдений, одинаково упорядоченных по обеим переменным.

N_d. [illegible]

Мультимедийные ресурсы

[illegible] Центра социологических ресурсов Видеорола.

[illegible]

На web-сайте компании вы сможете найти шестое издание книги *Statistics. A Too[illegible]* [illegible]

[illegible] можете получить доступ к разнообразным ресурсам, [illegible] для тех, кто изучает социологию. Зайдя в раздел Student Resources ("Ресурсы для студентов") на web-странице книги, вы найдете следующее:

- Обзор основ математики
- Обзор статистики
- Обучающие карты,
- Интернет-ссылки
- Дополнительные зада[illegible] к отдельным [illegible]
- Таблицы случайных чисел
- Примеры и упражнения с использованием пакетов [illegible] обработки данных MicroCase и SPSS
- [illegible] диаграммы "F[illegible]d the [illegible]"
- Проверка гипотез для переменных, измеряемых по порядковой шкале

Задачи

[illegible] 14.3 и 14.5 требуется рассчитать [illegible] для анализа тесноты и направления связи.

14.1. SOC [illegible] Штаты, опрос[illegible] тех, [illegible] английский язык не является родным. Вопрос: зависит ли привычка [illegible] к [illegible] США? Для каждой таб[illegible] а также определить силу и направление свя[illegible]

[illegible] в числе наблюдений в верхней левой ячейке на число наблюдений в нижней правой ячейке. [illegible] число наблюдений в верхней правой ячейке на число наблюдений в нижней левой ячейке.

а. В[illegible]

Степень владения английским языком	Длительность проживания в США: Менее пяти лет (низкая)	Более пяти лет (высокая)	Итого
Ни[illegible]			
[illegible]			
И[illegible]			

б. Совокупный доход семьи

Доход	Длительность проживания в США: Менее пяти лет (низкая)	Более пяти лет (высокая)	Итого
Ниже среднего по стране (1)	18	8	25
[illegible]			
Итого	25	25	50

в. Интенсивность контактов со страной происхождения

Контакты	Длительность проживания в США: Менее пяти лет (низкая)	Более пяти лет (высокая)	Итого
[illegible]			
Часто (2)	20	5	25
Итого	25	25	50

14.2. Для таблиц из задач 11.2, 11.6, 11.7, 11.9, 11.10 и 11.12 вычислить гамму. Поскольку эти таблицы построены для случайных выборок, требуется проверить рассчитанные значения коэффициента гамма на статистическую значимость.

14.3. Вычислить коэффициент гамма для таблицы, представленной в задаче 12.[illegible]. Если в задаче 13.2 для этой таблицы была рассчитана какая-нибудь номинальная мера связи, то теперь можно сравнить ее с гаммой. Вопросы: равны ли эти меры связи по величине? Характеризуют ли они тесноту связи одинаково? Какую информацию о связи дает гамма в отличие от номинальных мер связи?

14.4. SJ

Случайная выборка из 150 городов была ранжирована по количеству жителей (малые, средние и крупные города) и по уровню преступности (высокий и низкий). Вопрос: существует ли зависимость между размером города (по числу жителей) и уровнем преступности?

Уровень преступности	Размер города: Малые	Средние	Крупные	Итого
Низкий	21	7	8	46
Высокий	29	33	42	104
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

а. Охарактеризовать силу и направление связи.

б. Является ли данная связь статистически значимой?

14.5. SOC

[illegible] Некоторые исследования показывают, что в семьях по-разному относятся к [illegible] детей. В семьях среднего класса ро[illegible] следят за этим процессом и [illegible] его (к примеру, способствуют записи [illegible] их в детскую бейсбольную лигу Little League). С другой стороны, в рабочих [illegible]

[illegible] организовывают свой досуг самосто- [illegible]

Подтверждают ли приведенные ниже данные результаты этих исследований? Требуется в нескольких предложениях сформулировать свои выводы.

В детстве вы занимались организованным спортом или "дворовым"?	Белые воротнички	[illegible]
Организованным	15 [illegible]	[illegible]
Дворовым	10	[illegible]

14.6. PA

Возрастает ли с увеличением трудового стажа эффективность деятельности работников коммунальной сферы? Другими словами, действительно ли "старой гвардии" лучше неопытной "зеленой молодежи"? Муниципалитет одного города собрал данные об эффективности деятельности и продолжительное [illegible]

Эффективность [illegible]	[illegible]	[illegible]	Всего
Низкая	54	25	79
Средняя	149	254	[illegible]

[illegible] последующими показателя- [illegible] теста и рейтинг, отражающий показатели в работе, для случайной выборки из 75 [illegible]

[illegible]	[illegible]	[illegible]	Высокий	Итого
Среднее	9	8	7	24
Высокое	5	[illegible]	9	26
[illegible]	25	[illegible]	25	75

а. Существует ли связь между [illegible] переменными? Если да, то [illegible]

б. Является ли показатель гамма стати[illegible] [illegible] ровать силу и направление связи.

14.8. SW

[illegible]

Симптомы депрессии у ребенка	Уровень авторитаризма взрослых: Низкий	Средний	Высокий	Итого
Мало	7	8	9	24
Частично	15	10	18	43
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Итого	30	30	30	90

14.9. SOC

Существует ли зависимость между склонностью к предубеждениям и уровнем образования? В нескольких предложениях сформулировать свои выводы.

Склонность к предубеждениям	Уровень образования: [illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

14.10. SOC

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]
Довольн	[illegible]	[illegible]	[illegible]
Очень довольн	[illegible]	[illegible]	[illegible]
Итого	[illegible]	[illegible]	[illegible]

а. Охарактеризовать силу и направление [illegible]

б. Является ли данная связь статистически значимой?

14.11. [illegible]

Случайная выборка из 11 районов города [illegible] комплексный показатель [illegible] и материальных [illegible] исследования [illegible] более высокие значения переменных соответствуют более высокому качеству жизни и большей [illegible] жителей района.

	Качество жизни	Сплоченность жителей района
[illegible]	[illegible]	[illegible]

[illegible] *подхода пропорционального уменьшения ошибки.*

[illegible] мулировать выводы.

Несколько лет назад была начата программа профессиональной подготовки [illegible] подтверждается ли [illegible] Существует ли связь между исходными оценка [illegible]

[illegible]	[illegible] оценка	Фактическая оценка профессионализма
[illegible]	10	42

14.13 SOC

[illegible]

	[illegible]	[illegible]
F	60	[illegible]
J	39	126
K	29	7
M	15	48
[illegible]	5	4[illegible]
O	0	[illegible]

14.14 SOC

[illegible] против данной этнической группы. Вопросы: насколько близк [illegible]

[illegible]	Средняя оценка по шкале «Социальная удаленность»: Белые студенты	Темнокожие студенты
[illegible]	[illegible]	[illegible]
В. Американские индейцы	[illegible]	[illegible]

	2,8	3,5

14.15 [illegible]

[illegible] В данной задаче в качестве независимой переменной использован [illegible] ... [illegible] ... исследования 1996 года, которое в кратком виде приведено в приложении Ж.

а. Зависимость поддержки узаконенного права на аборт (имя переменной — ABANY) от годового дохода респондента:

Право на аборт?	Менее $24,9 тыс.	$24,9 тыс. — $50 тыс.	Более $50 тыс.	Итого
[illegible]	[illegible]	218	226	664
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Итого	[illegible]	517	476	1579

б. Зависимость поддержки смертной казни (имя переменной — CAPPUN) от годового дохода респондента:

Смертная казнь?	Менее $24,9 тыс.	$24,9 тыс. — $50 тыс.	Более $50 тыс.	Итого
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	837	757	712	2306

в. Зависимость поддержки [illegible] права [illegible] помощью врача (имя переменной — [illegible]) от дохода респондента

Право на [illegible]	Менее $24,9 тыс.	$24,9 тыс. — $50 тыс.	Более $50 тыс.	Итого
Да	343	34[illegible]	[illegible]	[illegible]

Половое воспитание в школе	Менее $24,9 тыс.	$24,9 тыс. — $50 тыс.	Более $50 тыс.	Итого
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

д. Зависимость [illegible]

[illegible]	Менее $24,9 тыс.	$24,9 тыс. — $50 тыс.	Более $50 тыс.	Итого
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Использование *SPSS for Windows* для вычисления порядковых мер связи

Демонстрационный пример с использованием SPSS 14.1. Изменяется ли отношение к сексу с возрастом? Другой подход к нахождению ответа

В разделе "Демонстрационный пример [illegible] связи между переменными *premarsx* (отношение респондента к интимной близости до брака) и преобразованной *age* (возраст респондента) были использованы [illegible] из таблицы. Теперь мы исследуем указанную зависимость с целью выяснить, даст ли нам она иную информацию. Для этого запустим процедуру Crosstabs (Таблицы сопряженности), причем в качестве независимой (столбцовой) переменной *age* (*возрастная категория*; это преобразованная переменная *age*). Если к данному моменту у читателя не сохранилась преобразованная версия переменной [illegible] ... [illegible] ционный пример с использованием SPSS 12.1" зависимой (строчной) переменной [illegible] ... по кнопке Statistics (Статистики) и запросить гамму. Также следует щелкнуть по кнопке Cells (Ячейки) и запросить процентные величины по столбцам. Щелкнуть

[illegible] кнопке ОК. При этом откроется окно [illegible] и выводятся следующие [illegible]

Зависимость отношения к добрачным интимным отношениям от возраста респондента

		AGER			
		[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible] предосудительно	Процент от суммы по столбцу	5,9%	8,1%	11,3%	8,4%
Лишь в некоторых [illegible]	Процент от суммы [illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible] предосудительного	Процент от суммы [illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	Процент от суммы по столбцу	100,0%	100,0%	100,0%	100,0%

Симметричные меры связи

		Значение	Асимптотическая стандартная ошибка	[illegible] (приблизительно)	Уровень значимости (приблизительно)
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]			

Показатель гамма, равный -0,27, указывает на умеренную по силе отрицательную связь. Респонденты более старшего возраста (значение 3 переменной *agecat*) проявили наибольшую нетерпимость к такому явлению, как интимная близость до брака: среди них процент давших ответ "Безусловно предосудительно" оказался максимальным. С другой стороны, более молодые [illegible] среди них процент давших ответ "В этом нет ничего предосудительного" оказался максимальным. С возрастом поддержка добрачного секса [illegible] выбирает ответ "В этом нет ничего предосудительного" [illegible]

Демонстрационный пример с использованием SPSS 14.2. Направление связи между порядковыми переменными

[illegible] связи между двумя порядковыми переменными [illegible] порядковым переменным — непростой вопрос, в [illegible] кодировки их переменных [illegible] меньшего количества измерения. В [illegible]

[illegible] Следует отметить, что для большинства интервально-относительных [illegible] могут указывать на меньшее количество, что вызвано произвольностью кодирования. В [illegible] [illegible] религиозных предубеждений может означать либо сильную предубежденность, либо полное отсутствие таковой.

Рассмотрим переменные *premarsx* и *agecat* из раздела "Демонстрационный пример с использованием SPSS 14.1". Отрицательное значение показателя гамма означает, что высокие значения одной переменной связаны с низкими значениями другой. Например, те респонденты, которые по переменной *agecat* получили значение 1 (наименьшее возможное), склонялись по переменной *premarsx* к ответу 4 (наибольшее возможное значение). Дело в том, что значение 4 переменной *premarsx* может представлять как "высокий уровень поддержки добрачного секса", так и "низкий уровень неприятия добрачного секса". Отрицательная гамма всегда указывает на обратную зависимость между ЗНАЧЕНИЯМИ переменных, однако при этом собственно [illegible] между переменными не обязательно должна быть отрицательной. Противоре[illegible] направления для порядковых переменных требует от исследователя внимательно рассмотреть таблицу, чтобы убедиться в правильности интерпретации направления [illegible]

Демонстрационный пример с использованием SPSS 14.3. Интерпретация направления связи: Действительно ли престижная работа приносит большее удовлетворение?

Какова связь между престижностью работы и удовлетворением от нее? На этот вопрос [illegible] могли бы ответить, что иметь престижную должность выгодно в [illegible] (не только с точки зрения более высокого дохода). В частности, [illegible] приносит человеку более выраженное чувство удовлетворенности своей работой. В то же время другие могли бы возразить, поскольку человек на менее престижной должности несет меньшую ответственность и испытывает меньшее психологическое давление, то он больше удовлетворен своей работой, чем тот, кто занимает престижную должность.

Для проверки [illegible] требуется запустить процедуру [illegible] для переменных *prestg80* (*престижность работы*) и *satjob* (*удовлетворенность работой*). Далее с помощью команды **Frequencies** (Частоты) найти медиану для [illegible] переменной *prestg80* и преобразовать ее в категориальную. Аналогичным образом [illegible]

[illegible]

[illegible] образовать в другую переменную") и присвоить преобразованной переменной [illegible] (например, *prestcat*). Схема перехода к новой переменной для [illegible]

0 thru 42 → [illegible]
43 thru 86 → 2

Для *satjob* значение 1 остается неизменным [illegible] 4 следует сгруппировать:

1 = 1

2 или 4 = 2

[illegible] выбран в качестве независимой [illegible]

[illegible]

[illegible]

[illegible]. Щелкнуть по кнопке **OK**, при этом открывается [illegible] экран выводятся следующие таблицы.

Зависимость удовлетворенности своей работой от ее престижности

		Престижность работы		
		1	2 [illegible]	Итого
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Итого	Частота наблюдений	551	555	1106

Критерии хи-квадрат

	Значение	df (число степеней свободы)	Асимптотическая значимость (двусторонний)	Точная значимость (двусторонний)	Точная значимость (односторонний)
Критерий хи-квадрат Пирсона	19.337	1	0.00[illegible]		
[illegible]					
Критерий линейной зависимости	19.3[illegible]9	1	0.000		
[illegible]					

[illegible] связи

		Значение	Асимптотическая стандартная ошибка	T (приблизительно)	Уровень значимости (приблизительно)
Связь между двумя порядковыми переменными	Тому	-0.260	[illegible]	[illegible]	[illegible]
Количество валидных наблюдений		1106			

Рассматриваемая [illegible] статистически значима (уровень значимости хи-квадрат меньше 0,05) и, на что указывает значение [illegible], равное -0.26, умеренна по силе и отрицательна по направлению. На основании этих данных [illegible]

[illegible] сформулировать ошибочный вывод: [illegible] по мере возрастания престижности работы удовлетворенность ею снижается. Ошибка появляется [illegible] рассмотреть [illegible] кодов для переменной *satjob*: *более высокие* значения [illegible] соответствуют *меньшему* уровню удовлетворенности. Поэтому "высокое" (2) значение переменной *prestg80* связано с "высоким" (1) значением переменной *satjob*. Несмотря на [illegible] связь является *положительной*: [illegible] удовлетворенность работой возрастает с увеличением ее престижности.

Упражнения

14.1. Повторить пример 14.[illegible], использовав в качестве независимых какие-нибудь другие порядковые переменные. Какие еще факторы могут влиять на отношение респондента к добрачному сексу? Среди прочих предлагается рассмотреть в качестве независимых переменные *degree* (*ученая степень*) или *class* (*социальный класс*). Ответить на вопрос: зависит ли отношение к добрачному сексу от уровня образования респондента или от того, к какому социальному классу он [illegible] принадлежит. В качестве независимых также можно рассмотреть переменные *income91*, *prestg80*, *educ* [illegible] предварительно требуется преобразовать *polviews* и *attend*.

14.2. Повторить пример 14.3, рассмотрев зависимость преобразованной переменной *satjob* от каких-нибудь других, отличных от *prestg80*, независимых переменных — например, от принадлежности к общественному классу (*class*) или ученой степени (*degree*). В нескольких предложениях охарактеризовать силу и направление рассмотренных связей. Ответить на вопрос: эти связи сильнее или слабее чем в случае *prestg80*? Особенно внимательно следует подходить к интерпретации направления связей.

14.3. Исследовать связь между преобразованными переменными *polviews* (*политические взгляды*) и *age* (*возраст*). Повышается ли с возрастом человека его консерватизм? Интерпретировать силу и направление этой связи.

15

Связь между переменными, измеренными по интервальной или относительной шкале

Цели изучения К концу этой главы вы смо[illegible]

1. Интерпретировать диаграмму рассеяния
2. [illegible]
3. [illegible] строить прямую регрессии методом наименьших квадратов интерпретировать ее и использовать для предсказания значений Y [illegible]
4. [illegible]
5. [illegible]
6. [illegible] мость

15.1 Введение

[illegible] тельной шкале[1]. Как будет показано, эти методы [illegible] как и прежде, интересуют три [illegible] Насколько сильна эта связь? Как [illegible]

[illegible]

15.2 Диаграммы рассеяния

[illegible]

Однако в случае анализа корреляционной связи между интервальными/относительными переменными первым шагом, как правило, будет построение и исследование диаграммы рассеяния. [illegible]

[illegible] иллюстрируем построение и интерпретацию диаграмм рассеяния на примере. Допу[illegible] интересует вопрос о том, как справляются с работой по дому в семьях, где работают оба супруга (так называемые "семьи с двойным доходом"). В частности [illegible] в таблице 15.1.

Таблица 15.1. Количество детей в семье и вклад мужа в выполнение работы по дому (вымышленные данные)

Семья	Количество детей	Количество часов, которые муж посвящает работе по дому
[illegible]	[illegible]	[illegible]
G	3	[illegible]
H	3	6
[illegible]	4	5
[illegible]	4	[illegible]

[illegible] и значения зависимой переменной Y — по вертикальной. Каждая точка на диаграмме [illegible]

…но 15.1). Семья А имеет по переменной X значение … [illegible] … уделяет работе по дому один час в неделю). Этому … [illegible] … диаграммы … [illegible]

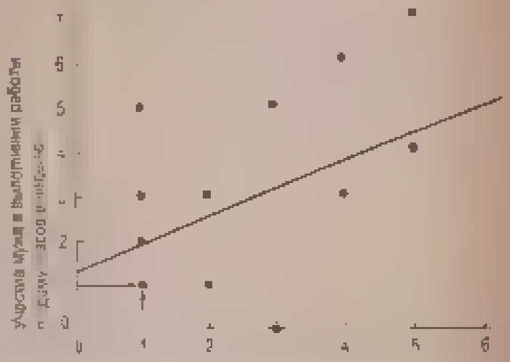


Рис. 15.1.
Зависимость участия мужа в выполнении работы по дому от количества детей в семье

Общий характер расположения точек (наблюдений) обобщает природу взаимосвязи между переменными. Наглядность характера расположения точек может быть повышена с помощью прямой линии, проведенной через скопление точек таким образом, чтобы она касалась каждой точки или проходила как можно ближе к ней. Метод подгонки линии к форме скопления точек будет рассмотрен в разделе 15.3. На данном этапе достаточно приближенный "на глаз". Обобщающая линия называется линией регрессии (см. рис. 15.1).

Даже приблизительно построенная … [illegible] … с их помощью можно … [illegible]

Чтобы удостовериться в существовании связи, необходимо вспомнить ее базовое … [illegible] … между двумя переменными существует связь, если … определения Y (зависимой переменной) … [illegible] … горизонтальной оси. Точки, расположенные … [illegible] … точки над каждым значением X (т.е. условные распределения … [illegible]

… [illegible] … линия регрессии … [illegible]

… [illegible]

… [illegible] линии регрессии. На рис. 15.1 изображена положительная связь: с увеличением X (количества детей) возрастает и Y (участие мужа в работе по дому). Значит, в семьях с большим количеством детей мужья активнее участвуют в выполнении работы по дому. В случае отрицательной связи прямая регрессии была бы наклонена в противоположном направлении, при этом высокие значения одной переменной связаны с низкими другой.

На рис. 15.2 представлены примеры абсолютной положительной, абсолютной отрицательной и нулевой связи (т.е. абсолютное отсутствие зависимости) между двумя переменными.

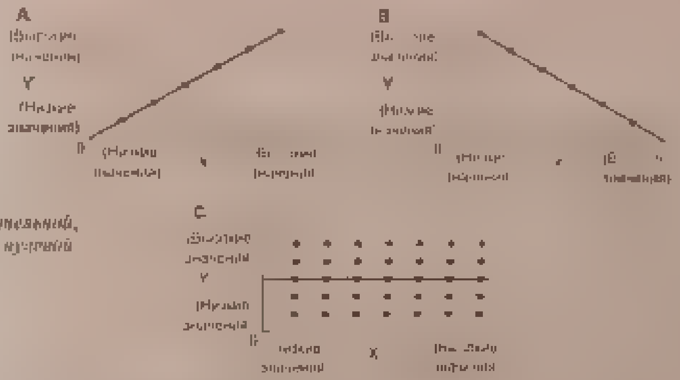
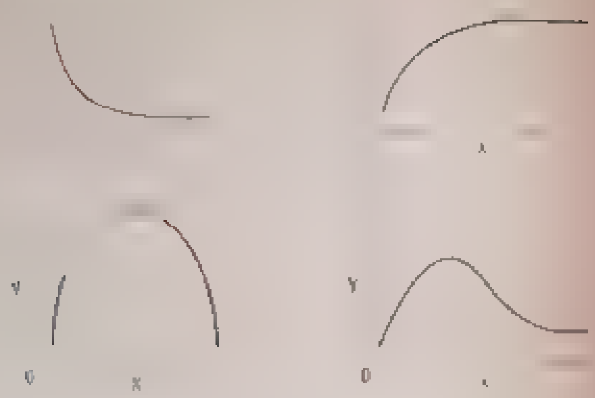


Рис. 15.2.
Примеры положительной, отрицательной и нулевой связи

Ключевое допущение, на котором основаны рассматриваемые в данной главе статистические методы, состоит в том, что между переменными существует линейная связь. Другими словами, точки диаграммы рассеяния должны образовывать скопление, которое можно аппроксимировать прямой линией. Значительные отклонения от линейности требуют использования статистических методов, выходящих за рамки данной главы. Тем не менее, примеры нескольких криволинейных зависимостей представлены на рис. 15.3. Если диаграмма рассеяния указывает на существование между переменными нелинейной связи, то изложенные в этой главе методы следует применять с большой осторожностью или вообще их не использовать. Проверка связи на линейность — едва ли не самая важная причина построения хотя бы приблизительной диаграммы рассеяния перед тем, как переходить к статистическому анализу. Если обнаружится, что исследуемая зависимость нелинейная, то переменные следует рассматривать как порядковые, а не интервальные/относительные



15.3. Регрессия и предсказание

...семье шесть детей. В рассмотренной выборке отсутствуют семьи с ... детьми, и если продолжить линию регрессии настолько, чтобы включить ... (см. рис. 15.4).

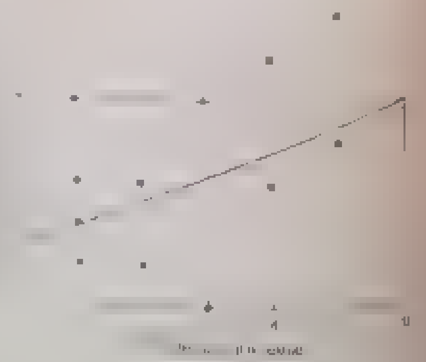


Рис. 15.4

...от фактического ...

...проводится вертикальная прямая...

...регрессии, от которой, в свою очередь, проводится ... до пересечения с осью Y. Эта точка и является Y'. Мы можем ... что в семье с шестью детьми муж уделяет работе по дому приблизительно ... часов в неделю.

Конечно, ... может оказаться ...

...регрессии ... построить такую линию регрессии, которая максимально ... точек, отражающих ... наблюдения ... самым лучшим образом описывает ... между переменными. Существует ли вообще такой "наилучший" ... И если есть, то как ее определить?

Проводя линию регрессии "на глаз", мы стремимся к тому, чтобы она ... проходила как можно ближе к ним. Мы уже знаем, что точки для каждого значения X могут рассматриваться как условные распределения ... переменной Y. Внутри каждого условного распределения переменной Y ... точка, вокруг которой ... Ранее среднее ... распределения было определено как точка, вокруг которой ...

$$\sum (Y - \dots)$$

Следовательно, если линия регрессии ... проходить ... каждого условного среднего Y, то мы получим прямую, проходящую максимально близко от всех точек...

Условное среднее находим суммированием всех значений Y ... каждому значению X, и последующим делением ... на число наблюдений ...

...работе по дому соответственно ..., 2, 3 и 5 часов в неделю. Имеем X = 1, Y = 1, 2, 3, 5, условное среднее Y для X = 1 равно 2,75 (11 : 4 = 2,75). Значит, мужья в семьях с одним ребенком занимаются работой по дому в среднем 2,75 часа в неделю. Аналогично вычисляем условные средние Y для всех остальных значений X. Результаты ... сведены в табл. 15.2 и графически отображены на рис. 15.5.

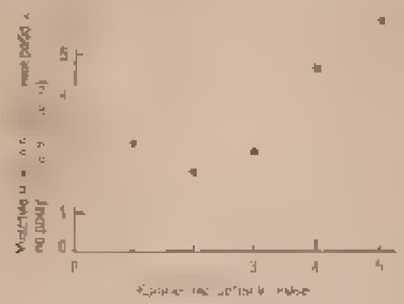
Таблица 15.2. Условные средние Y (участие мужей в работе по дому) для различных значений X (количество детей в семье)

Количество детей (X)	Количество часов, которые муж посвящает работе по дому (Y)	...
[illegible]	[illegible]	[illegible]

Вспомним, для чего ...

[illegible] ...точек насаждения. [illegible] ...к каждой точке и потому она — лучше всего подогнанная линия регрессии.

Рис. 15.5 [illegible]



[illegible] проведем [illegible] через точки на рис. 15.5 (эти точки — условные средние Y), [illegible] определению быть лучше всего подогнанной прямой регрессии. Однако видно, что фактически эта линия — не прямая. В действительности условные средние попадают на идеальную прямую линию очень редко, при существовании между переменными X и Y абсолютной связи. Тем не менее, мы должны выдержать условие линейности. Поэтому изменим наш критерий и определим прямую регрессии как уникальную прямую линию, которая касается всех условных средних Y или проходит как можно ближе к ним. Формула 15.1 — это уравнение прямой регрессии, построенной по методу наименьших квадратов и лучше всего подогнанной к скоплению [illegible].

Формула 15.1 $$Y = a + bX$$

где: Y — значение зависимой переменной,
a — свободный член (отрезок, отсекаемый на оси Y, или точка пересечения прямой регрессии с осью Y),
b — коэффициент наклона (выражает наклон линии регрессии [illegible] изменение Y при единичном изменении X),
X — значение независимой переменной.

[illegible] свободный член (a), [illegible] вертикальной осью Y. Во-вторых, коэффициент наклона b [illegible] по методу наименьших квадратов [illegible] наименьших квадратов") — это [illegible] переменной Y при изменении независимой [illegible] можем рассматривать как меру воздействия независимой X на переменную Y. Если между переменными [illegible] изменения значе[illegible] Y, поэтому коэфф[illegible]. Чем меньше [illegible]

[illegible] X воздействует на переменную Y (т.е. чем слабее связь между этими двумя переменными), тем меньшее значение имеет коэффициент наклона b. При отсутствии связи между рассматриваемыми переменными прямая регрессии, построенная по методу наименьших квадратов, параллельна оси X, а b равно 0.00 (т.е. прямая не имеет наклона).

[illegible] квадратов (см. формулу 15.1), — это единственная прямая, [illegible] данным лучше всего [illegible] максимально близко ко всем условным средним Y.

Прежде чем изучить способ предсказания Y [illegible] *математического основания методики применения прямой регрессии для прогнозирования значений Y по значениям X читатели могут рассмотреть задачи 15.1, 15.2, 15.3 и 15.5.*

15.4. Вычисление *a* и *b*

Поскольку величина b нужна для определения a, то вначале вычислим наклон прямой, построенной по методу наименьших квадратов (см. формулу 15.2).

Формула 15.2 $$b = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2}$$

Числитель этой формулы называется *ковариацией* X и Y. Это мера совместного изменения X и Y, а ее величина отражает направление и силу связи. На практике, однако, использование формулы 15.2 утомительно, поэтому рекомендуется применять производную от нее формулу 15.3.

Формула 15.3 $$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$$

где: b — коэффициент наклона,
N — число наблюдений,
$\sum XY$ — сумма перекрестных произведений значений,
$\sum X$ — сумма значений X,
$\sum Y$ — сумма значений Y,
$\sum X^2$ — сумма квадратов X.

На первый взгляд формула 15.3 выглядит громоздкой. Однако процесс вычислений упрощается, если его организовать в табличной форме (см. таблицу 15.3), где каждой из четырех величин, необходимых для получения b, отведен отдельный столбец. Данные взяты из рассмотренного выше примера семей с двойным доходом (таблица 15.1).

Таблица 15.3 [illegible] b

4			4
9		25	16
5		9	12
[illegible]		49	35
30	112	184	125

$\Sigma XY = 125$

[illegible]

В первых двух столбцах таблицы 15.3 приведены исходные значения X и Y для каждого наблюдения. В третьем столбце приведены квадраты X, в четвертом — квадраты Y, и в пятом — перекрестные произведения значений X и Y. Подставив в формулу [illegible] соответствующие [illegible]

(1500 – 280)

(6344 – 1024)

$b = 0.69$

Коэффициент наклона 0.69 означает, что при каждом единичном [illegible] значение Y возрастает на 0.69 единицы. В [illegible]

[illegible] единичное увеличение переменной X) [illegible]

будет уделять работе по дому на 0.69 часа в неделю [illegible]

[illegible]

Формула 15.4

[illegible]

$$a = \bar{Y} - b\bar{X}$$

$a = 3.49$ [illegible]

$a = 1.49$

[illegible] секает вертикальную ось в точке, где Y равен 1.49.

Формула линии регрессии, построенной по методу наименьших квадратов, для наших выборочных данных имеет вид

$$Y' = 1.49 + 0.69X$$

Эта формула может использоваться для оценки или предсказания значения Y для любого значения X. В разделе 15.3 с помощью линии регрессии, построенной «на глаз», было предсказано значение Y [illegible] с шестью детьми ($X = 6$) — приблизительно [illegible]. Теперь мы можем [illegible], насколько точным было наше [illegible] предсказание.

$$Y' = a + bX$$

[illegible]

$$Y' = 5.63$$

С помощью линии регрессии, построенной по методу наименьших квадратов, мы можем предсказать, что в семье с шестью детьми, в которой работают оба супруга, муж посвящает работе по дому 5.63 часа в неделю. [illegible] рекомендуется самостоятельно определить, сколько часов в неделю будет посвящать работе по дому муж в семье с семью детьми ($X = 7$).

Следует заметить, что даже такие, полученные по методу наименьших квадратов [illegible]

Y маловероятно, за исключением того редкого случая, когда [illegible] абсолютна и абсолютно линейна. Точность предсказаний повышается с увеличением [illegible]. Поскольку при [illegible] зависимости между переменными группирование точек вокруг прямой регрессии, построенной по методу наименьших квадратов, больше [illegible]

Для практики коэффициент наклона и свободный член [illegible] в конце данной главы, однако особенно рекомендуются решения 15.1 – 15.5. [illegible]

[illegible] небольшое количество исходных данных, что облегчает освоение методики вычислений.

15.5. Коэффициент корреляции (r Пирсона)

В разделе 15.4 говорилось, что наклон линии наименьших квадратов (b) является мерой воздействия X на Y. Поскольку наклон — это изменение Y при единичном изменении X, то величина b возрастает с усилением связи между переменными. Однако диапазон изменения b — не от 0 до 1, поэтому использовать этот параметр в качестве меры связи не удобно. При проведении исследований широко (если не исключительно) для измерения связи между интервально-относительными переменными применяется статистика, называемая **r Пирсона**, или коэффициент корреляции. Подобно описанным в главе 14 порядковым мерам связи, r Пирсона изменяется в диапазоне от 0 до ±1,00, где 0,00 указывает на отсутствие связи, а +1,00 и −1,00 — на абсолютную положительную или отрицательную связь соответственно. r Пирсона определяют по формуле:

Формула 15.5

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\left[\sum (X - \bar{X})^2\right]\left[\sum (Y - \bar{Y})^2\right]}}$$

[illegible] *вариации X и Y*. На практике, однако, использовать формулу 15.5 [illegible], поэтому рекомендуется применять производную от нее формулу 15.6.

Формула 15.6

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\left[N\sum X^2 - (\sum X)^2\right]\left[N\sum Y^2 - (\sum Y)^2\right]}}$$

[illegible] подобной таблице 15.3. Все необходимые члены уравнения 15.6 уже имеются в таблице 15.3, поэтому подставим их.

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\left[N\sum X^2 - (\sum X)^2\right]\left[N\sum Y^2 - (\sum Y)^2\right]}}$$

$r =$ [illegible]

$r =$ [illegible]

$r =$ [illegible]

$r = 0,50$

Значение r, равное 0,50, указывает на умеренную по силе положительную линейную связь между переменными [illegible] *данных с целью упрощения вычисления r Пирсона. Рекомендуем [illegible] имеющих небольшое количество данных (в частности с 15.1), поскольку вычисления в них расчеты проще.*

15.6. Интерпретация коэффициента корреляции: коэффициент детерминации r^2

Коэффициент корреляции r Пирсона — это показатель тесноты линейной связи между переменными. Если значение 0,00 говорит о полном отсутствии связи, а ±1,00 — об абсолютной линейной связи, то промежуточные значения не имеют прямой интерпретации. Конечно, можно описывать связь, исходя из того, насколько близко значение r к какому-либо крайнему значению (0,00 или ±1,00). Например, если r по величине близко к 0,00, то связь можно охарактеризовать как "слабую", а в случае близости значения r к ±1,00 — как "сильную". Однако такая оценка несколько субъективна.

К счастью, можно добиться более объективной интерпретации. Для этого требуется рассчитать дополнительную статистику, называемую **коэффициентом детерминации**. Она представляет собой просто квадрат r Пирсона (r^2), а логика ее интерпретации подобна подходу пропорционального уменьшения ошибки. Вспомним, что такой подход для мер связи — это предсказание значения зависимой переменной при двух разных условиях. Сначала мы предсказываем значение Y, игнорируя информацию об X, затем выполняем предсказание уже с учетом этой информации. Для r используются несколько другие методы предсказания и построения конечной статистики, поэтому требуется введение ряда новых понятий.

При работе с переменными, принадлежащими к интервальной/относительной шкале, предсказание Y при первом условии (игнорирование X) — это среднее значений Y (т.е. $\bar{Y}$) для каждого наблюдения. Если информация об X отсутствует, то данная стратегия предсказания оптимальна, поскольку мы знаем, что среднее любого распределения ближе ко всем его значениям, чем любая другая точка этого распределения. Вспомним введенный в главе 3 принцип [illegible] и отражающую его формулу

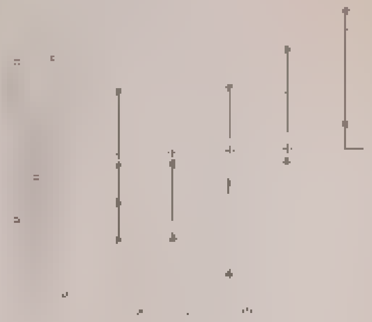
$$\sum (Y - \bar{Y})^2 = \text{[illegible]}$$

Значения любой переменной отклоняются от среднего значения меньше, чем от любой другой точки. Следовательно, если бы нам удалось предсказать среднее Y для каждого наблюдения, то мы сделали бы меньше ошибок предсказания, чем в случае предсказания любого другого значения Y.

Конечно, даже при добросовестном выполнении описанной стратегии будет допущено немало ошибок предсказания. Количество ошибок представлено на рис. 15.6, который отражает связь между количеством детей в семье и участием мужа в работе по дому. Также на рисунке показано среднее Y ($\bar{Y}$). Вертикальные отрезки от фак

тических значений до предсказанного (т.е. $\bar{Y}$) соответствуют величине ошибки, которая была бы сделана в случае предсказания без учета X.

Рис. 15.6 (Предсказание [illegible] с двойным доходом)

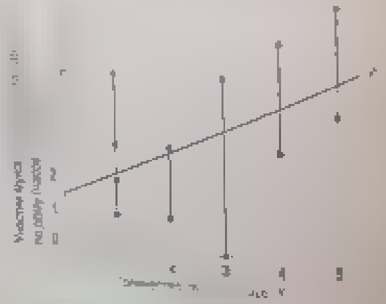


Определить величину ошибки предсказания при первом условии (игнорирование X) можно вычитанием среднего Y из каждого фактического значения Y с последующим возведением полученных отклонений в квадрат и суммированием их. Итоговое число — $\sum(Y - \bar{Y})^2$ — называется **общей вариацией** переменной Y. Теперь мы имеем визуальное представление (рис. 15.6) и метод вычисления ошибки предсказания Y в случае игнорирования X. Однако ниже будет показано, что для нахождения величины коэффициента детерминации r^2 нам на самом деле не требуется вычислять [illegible]

Следующий шаг — определить, насколько информация об X улучшает наше предсказание значения Y. Если эти две переменные связаны линейно, то предсказание значений Y с помощью уравнения регрессии, построенного по методу наименьших квадратов, будет выполняться с учетом X, что уменьшит ошибки предсказания. Итак, при втором условии предсказанное значение Y для каждого значения X равно [illegible]

На рис. 15.7 представлены данные о семьях с двойным доходом, а также линия [illegible] соответствуют величине ошибки, которая осталась после того, как мы сделали предсказания с учетом X.

Рис. 15.7 [illegible]



Теперь можно точно определить уменьшение ошибки, которое стало результатом учета X: найти две различные суммы, сравнить их с общей вариацией Y и получить статистику-показатель улучшения прогнозирования.

Первая сумма, называемая **объясненной вариацией**, [illegible] способности предсказывать Y при учете X. Эту сумму получают вычитанием $\bar{Y}$ (предсказанного значения Y без учета X) из значения, предсказанного уравнением регрессии (Y' или значение Y, предсказанное с учетом сведений о X) для каждого наблюдения, с последующим возведением разностей в квадрат и суммированием. Изложенные операции можно обобщить выражением $\sum(Y' - \bar{Y})^2$. Полученное число сравнивается с общей вариацией Y для оценки того, насколько знание об X улучшило нашу способность предсказывать Y. Математически сказанное выражается следующим образом:

Формула 15.7

$$r^2 = \frac{\sum(Y' - \bar{Y})^2}{\sum(Y - \bar{Y})^2} = \frac{\text{Объясненная вариация}}{\text{[illegible]}}$$

Значит, коэффициент детерминации, или r^2, — это доля общей вариации Y, объясненной переменной X. Подобно другим мерам пропорционального уменьшения ошибки, r^2 точно показывает, насколько X помогает улучшить предсказание (понимание и объяснение) Y.

Практический пример 15.1

По пяти городам была собрана информация о количестве случаев [illegible] волнений (беспорядки, забастовки и т.д.) за последний год и об [illegible] тицы. Данные представлены в таблице. Вопрос: связаны ли между собой эти переменные?

Город	Уровень безработицы (X)	Гражданские волнения (Y)	X^2	Y^2	XY
A	22	25	484	625	550
B	20	13	400	169	260
C	10	10	100	100	100
D	15	5	225	25	75
E	9	0	81	0	0
	76	53	1290	919	985

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$$

$$b = \frac{(5)(985) - (76)(53)}{(5)(1290) - (76)^2}$$

$$b = \frac{[illegible]}{[illegible]}$$

$$b = [illegible]33$$

[illegible] что на каждое единичное изменение X (т.е. [illegible] уровня безработицы [illegible] приходится изменение Y [illegible] гражданских волнений [illegible]

Свободный член a:

$$a = \bar{Y} - b\bar{X}$$

$$a = \frac{[illegible]}{5} - [illegible]$$

$$a = 10{,}6 - 1{,}33 \times 15{,}20$$

$$a = [illegible]$$

$$a = [illegible]$$

Уравнение регрессии [illegible]

$$\hat{Y} = a + bX = [illegible] + [illegible]33X$$

Коэффициент корреляции

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

$$r = \frac{(5)(965) - (76)(53)}{\sqrt{[[illegible] - (76)^2][[illegible] - (53)^2]}}$$

$$r = \frac{897}{\sqrt{(67[illegible])([illegible])}}$$

$$r = \frac{897}{\sqrt{[illegible]}}$$

$$r = \frac{897}{1097{,}16}$$

$$r = 0{,}82$$

Можно сделать вывод, что между переменными существует сильная положительная связь. С возрастанием уровня безработицы [illegible] и количество [illegible] Коэффициент детерминации r^2 равен $(0{,}82)^2$, или [illegible] вариации гражданских волнений объясняются уровнем безрабо[illegible]

Выше мы назвали улучшение предсказания Y при помощи X объясненной вариацией. Из смысла этого термина следует, что некоторая часть вариации Y будет "необъясненной" или не обусловленной переменной X. На рис. 15.7 вертикальные отрезки представляют **необъясненную дисперсию** или разницу между нашим лучшим предсказанием Y с учетом X и фактическими значениями.

[illegible]о, необъясненная вариация — это рассеяние фактических значений вокруг прямой регрессии. Ее можно получить вычитанием предсказанных значений Y из фактических для каждого наблюдения с последующим возведением разностей в квадрат и суммированием. Последовательность вычислений выражается формулой $\sum(Y - \hat{Y})^2$, а результирующая сумма является величиной ошибки, которая осталась даже после того, как мы сделали предсказания с учетом X. Часть полной вариации Y, необъясненной переменной X, можно определить вычитанием величины r^2 из 1,00. Необъясненная вариация обычно обусловлена влиянием комбинации других переменных, ошибкой измерения и фактором случайности.

Объясненная и необъясненная вариации взаимосвязаны — когда одна из этих величин возрастает, другая уменьшается. Кроме того, чем сильнее линейная связь между X и Y, тем больше величина объясненной вариации и меньше — необъясненной. В случае абсолютной связи ($r = \pm1{,}00$) необъясненная вариация будет равна 0, а $r^2 = 1{,}00$. Это указывает на то, что X объясняет или обусловливает все изменения Y, т.е. что мы, зная X, можем безошибочно предсказать Y. С другой стороны, если между X и Y отсутствует линейная связь ($r = 0{,}00$), то объясненная вариация будет равна 0, а $r^2 = 0{,}00$. В таком случае мы делаем вывод, что X вообще не объясняет вариации Y и не улучшает нашу способность предсказывать Y.

[illegible]ия между двумя этими крайними величинами показывают, насколько знание о переменной X повышает нашу способность предсказывать или объяснять значения переменной Y. Мы выяснили, что для семей с двойным доходом $r = 0{,}50$. Квадратом этой величины будет коэффициент детерминации ($r^2 = 0{,}25$), который говорит о том, что количество детей в семье (X) объясняет 25% полной вариации участия мужа в выполнении работы по дому (Y). Предсказывая количество часов, которые муж посвящает работе по дому, мы, если наш прогноз будет основываться на количестве детей в семье и выполняться по прямой регрессии, сделаем на 25% меньше ошибок по сравнению с предсказанием среднего Y для каждого наблюдения, игно[illegible] вывод, 75% вариации Y не объясняются переменной X, предположительно из-за ком[illegible] факторов [illegible]ности. *Для практического освоения методики интерпретации r^2 читатели могут рассмотреть любую задачу в конце данной главы.*

15.7 Проверка коэффициента корреляции r Пирсона на значимость

[illegible]

пользуется $r = 0{,}50$ [illegible]

[illegible] двумерное нормальное распределение [illegible]

[illegible] правилом: если точки образуют достаточно симметричное сигарообразное [illegible]

Шаг 1. Принятие предположений

[illegible]

Переменные измерены по интервальной/относительной шкале

Двумерные нормальные распределения

[illegible]

[illegible] однородность дисперсии,

Нормальное распределение выборочных статистик

Шаг 2. Формулирование нулевой гипотезы

H_0: [illegible]

[illegible]

Шаг 3. Подбор распределения выборочных статистик и задание критической области

Поскольку за нулевую гипотезу [illegible] отсутствия зависимости [illegible], то распределение всех возможных выборочных [illegible]

коэффициентов корреляции аппроксимируется t-распределением. Число степеней свободы равно $(N - 2)$.

Распределение выборочных статистик = t-распределение,

Альфа = [illegible]

Число степеней свободы = N [illegible]

[illegible]

Шаг 4. Вычисление статистики критерия

$$t\ (\text{экспериментальное}) = r\sqrt{\frac{N [illegible]}{[illegible]}}$$

[illegible] (0,50) [illegible]

$$t\ (\text{экспериментальное}) = (0{,}50)\sqrt{\frac{10}{0{,}75}}$$

[illegible]

t (экспериментальное) = (0,50)(3,65)

t (экспериментальное) = 1,83

Шаг 5. Принятие решения

Поскольку статистика критерия не попадает в критическую область, граница которой задается t (критическим), то нулевую гипотезу отклонить нельзя. Даже несмотря на существование тесной связи между переменными в выборке, мы не получили достаточных подтверждений тому, что эти переменные связаны между собой и в генеральной совокупности. Проверка показала, что если нулевая гипотеза верна в генеральной совокупности переменные не связаны между собой, то выборочное $r = 0{,}50$ могло получиться по чистой случайности. *Для практического освоения методики проверки значимости коэффициента корреляции r Пирсона и интерпретации результатов читателям рекомендуется рассмотреть задачи 15.1, 15.2, 15.4, 15.6, 15.8 и 15.9.*

15.8. Интерпретация статистик: действительно ли в экономике США господствуют тесно взаимосвязанные корпорации?

[illegible]

[illegible] связаны между собой через лиц, которые входят в два или больше разных правлений (такие связи получили название "взаимосвязанные директораты"), то можно [illegible] признать, что в экономике США господствует малочисленная [illegible] элита. Дополнительным доказательством [illegible] была бы и обнаруженная положительная корреляция между размером корпорации и количеством ее связей с другими корпорациями, поскольку было бы установлено, что самые могущественные субъекты рынка имеют и наибольшее количество взаимосвязей с другими субъектами, что позволяет им осуществлять максимальный контроль над экономикой. Следовательно, требуется ответить на такие вопросы: имеет ли значение размер корпорации? Являются ли самые важные корпорации более всего [illegible] с другими корпорациями? В данном разделе будет исследована связь между количеством взаимосвязанных директоратов, которые имеет одна корпорация, и ее доходами. Использована случайная выборка из 125 корпораций США за 1995 год[3].

Компании, включенные в выборку, представляют различные отрасли экономики, в том числе транспорт, финансы, коммунальные услуги. В корпорациях бывает до 30 директоров, поэтому вполне вероятно, что между двумя корпорациями может быть много связей. В качестве зависимой переменной примем общее число связей, которые отдельная корпорация имеет с другими корпорациями. Независимая переменная — это общий доход корпорации (в миллиардах долларов) за 1995 год. В таблице 15.4 представлены основные описательные статистики для этих двух переменных.

Таблица 15.4. Основные описательные статистики

Переменная	N (число наблюдений)	Среднее	Среднеквадратичное отклонение
Общее число связей одной корпорации	125	6,9920	5,7788
Доходы за 1995 год в млрд долл	125	15,6259	20,4317

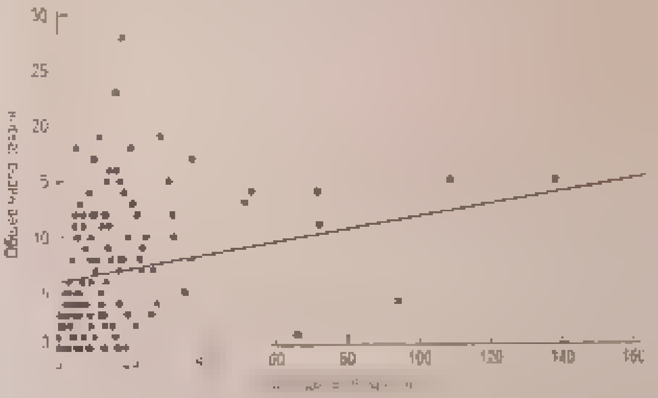
Первый шаг процесса оценки связи между интервальными [illegible] переменными — построение диаграммы рассеяния (рис. 15.8). Общее число связей каждой корпорации отложено по вертикальной оси Y, а доходы — по горизонтальной оси X. В первую очередь в этой диаграмме рассеяния привлекает внимание отсутствие хрестоматийной зависимости между переменными. Читатели должны четко представлять, что собой представляет идеальная или абсолютная положительная или отрицательная связь, но в то же время следует отдавать себе отчет в том, что такие [illegible] диаграммы рассеяния — редкое явление. В ходе реального исследования более вероятно получить размытую парную связь, как в данном примере.

Какие выводы можно сделать из рассмотрения этой диаграммы? Линия регрессии не горизонтальна, значит между переменными все-таки существует связь. Учитывая отсутствие плотного скопления точек вдоль прямой регрессии, заключаем, что эта связь недостаточно сильна. Подъем прямой слева направо говорит о положительной

[3] Данные взяты из доклада Barnes, Roy C. "Patterns of Corporate Interlocking in the United States 1962—1995". Статья представлена на ежегодной конференции Южного социологического общества. [illegible], штат Луизиана.

[illegible] связи: корпорации с более высокими доходами действительно имеют больше связей с другими корпорациями. Отметим также, что большинство точек наблюдений находятся в области малых значений переменной X (т.е. это корпорации с меньшими доходами) и они характеризуются большим разнообразием количества тех директоров, которые занимают пост одновременно в нескольких корпорациях. Так, отдельные корпорации с низкими доходами имели всего несколько "директоров-совместителей" (или вообще ни одного), тогда как другие из этой же категории (с низкими доходами) имели много таких связей с другими корпорациями. При таком симметричном распределении можно было бы говорить об отсутствии связи между переменными. Однако за счет более крупных (в смысле, с большими доходами) корпораций, среди которых наблюдается большее количество взаимосвязанных директоратов, связь в целом получилась положительной.

Рис. 15.8 [illegible]



Второй шаг процесса оценки зависимости — получить уравнение линии регрессии для всех наблюдений. Мы пропустим промежуточные действия, необходимые для получения коэффициентов регрессии (a и b) и сразу приведем уравнение

[illegible]

[illegible]

Линия регрессии пересекает ось Y в точке с координатой $Y = 5,52$, и каждое увеличение переменной X (дохода корпорации) на единицу приводит к увеличению количества взаимосвязей данной корпорации с другими на 0,14.

[illegible]

которая говорит, что между доходами и количеством взаимосвязей между корпорациями существует, в лучшем случае, умеренная связь. Проверка на значимость [illegible] не забудем, что до настоящего момента мы рассматривали не генеральную совокупность, [illegible] выборку из 125 корпораций США [illegible]

[illegible]

Мы отклоняем нулевую гипотезу об "отсутствии связи" и делаем вывод о существо- [illegible]

[illegible] означает, что уровень дохода корпорации объясняет только 11% вариации количе- [illegible]

[illegible] сти общего количества взаимосвязей, которые одна корпорация имеет с другими через своих директоров. Эти результаты подтверждают мнение о большей [illegible]

Заключение

Эти выводы основаны на примере, который рассматривался в данной главе.

1. Мы начали с вопроса: существует ли связь между количеством детей в семьях с двойным доходом и количеством часов, которые муж посвящает работе по дому? Наблюдения были представлены на диаграмме рассеяния (рис. 15.1), рас[illegible]ние которой привело к мнению, что между переменными существует положительная связь. Точки диаграммы рассеяния образовали скопление, которое можно аппроксимировать прямой линией, — значит, связь между переменными прибли[illegible]

2. Значения переменной Y можно предсказать при помощи построенной "на глаз" прямой регрессии, однако большая точность достигается в случае использования прямой, построенной по методу [illegible] наименьших квадратов — это прямая, наилучшим образом подогнанная под данные путем минимизации вариации Y. Используя формулу прямой, полученной методом наименьших квадратов ([illegible] + bX), мы нашли наклон прямой b, равный 0,69 и означающий, что каждый дополнительный ребенок в семье (единичное изменение переменной X) приводит к увеличению участия мужа в работе по дому на 0,69 часа в неделю. На основании этой формулы мы также предсказали, [illegible] шестью детьми ($X = 6$) мужья посвящают [illegible]

3. Коэффициент корреляции r Пи[illegible] тем тесноты общей линейной с[illegible] ($r^2 = 0,25$), показывающий, что 25% общей вариации Y (участия мужа в выполнении работы по дому) объясняется [illegible] количеством детей в семье).

4. Поскольку рассмотренные нами 12 семей составляют случайную выборку, мы проверили коэффициент корреляции r Пирсона на статистическую значимость и обнаружили, что, при уровне значимости 0,05, мы не можем утверждать о существовании линейной связи между этими переменными в генеральной совокуп[illegible]

5. Мы получили много информации о данной парной связи. Мы знаем ее силу и направление, а также уравнение прямой регрессии, которая наилучшим образом отражает влияние переменной X на Y. Мы знаем, какого изменения Y нам следует ожидать при единичном изменении X. В общем, об этой связи между интервально-относительными переменными мы имеем больше сведений, к тому же [illegible]ных, чем когда-либо получали о связях между порядковыми и номинальными переменными. Это вполне объяснимо, поскольку данные, являющиеся результатом измерений по интервальной/относительной шкале, точнее и гибче [illegible], которые получены в результате измерений по порядковой и номинальной [illegible]

Сводка формул

[illegible] ия регрессии, построенная методом наименьших квадратов

15.1 $$Y = a + bX$$

Формула для определения коэффициента наклона

$$\sum(X - \bar{X})^2$$ [illegible]

[illegible]

15.3 $$b = \frac{[illegible]}{N\sum X^2 - (\sum X)^2}$$

Свободный член

15.4 $$a = \bar{Y} - b\bar{X}$$

Формула определения r Пирсона

15.5 $$r = [illegible]$$

Формула для вычисления r Пирсона

15.6 $$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][illegible]}}$$

Коэффициент детерминации

$$r^2 = \frac{\sum(Y' - \bar{Y})^2}{\sum(Y - \bar{Y})^2} = \frac{\text{Объясненная вариация}}{\text{Общая вариация}}$$

Словарь

Гомоскедастичность. Допущение статистической модели об дисперсии значений Y для всех значений X, которое принимается при проведении проверки значимости r Пирсона.

Двумерное нормальное распределение. Допущение статистической модели о нормальном распределении обеих переменных, которое принимается при проведении проверки r Пирсона на значимость.

Диаграмма рассеяния. Графическое изображение взаимосвязи между двумя переменными.

Коэффициент детерминации (r^2). Доля всей дисперсии Y, которая объясняется переменной X. Находится возведением в квадрат коэффициента корреляции (r Пирсона).

Коэффициент корреляции r Пирсона. Мера [illegible] переменных, измеряемых по интервально-относительной шкале. ρ (греческая буква ро) — это символ для r Пирсона в генеральной совокупности.

Коэффициент наклона (b). [illegible] b — это [illegible] наклона прямой регрессии.

Линейная связь. Связь между двумя переменными, при которой совокупность точек наблюдений на диаграмме рассеяния можно аппроксимировать прямой [illegible]

Линия регрессии. Это единственная [illegible], которая обобщает взаимосвязь между двумя переменными. [illegible] регрессии к точкам данных выполняется с [illegible] квадратов, благодаря [illegible] на максимальном расстоянии.

Необъясненная вариация. Доля общей вариации Y, которая не объясняется [illegible] переменной X. Находится по формуле $\sum(Y - Y')^2$.

Общая вариация. Разброс значений Y вокруг среднего $\bar{Y}$. Находится по формуле $\sum(Y - \bar{Y})^2$.

Объясненная вариация. Доля общей вариации Y, которая объясняется [illegible] переменной X. Находится по формуле $\sum(Y' - \bar{Y})^2$.

Свободный член (a). Точка пересечения прямой регрессии [illegible]

Условное среднее Y. Среднее всех значений Y для каждого [illegible] значения X [illegible] предсказанного значения Y.

Мультимедийные ресурсы

The [illegible] Society (Виртуальное [illegible] ресурсов Вадсворта)

http://[illegible].com

На web-сайте компании вы сможете найти шестое издание книги *Statistics: A Tool for Social Research* ("Статистика как инструмент социологических исследований"), а также получите доступ к разнообразным ресурсам, полезным для тех, кто изучает [illegible] Student Resources ("Ресурсы для студентов") на web-странице книги, вы найдете следующее.

- Обзор основ математики,
- Обзор статистики
- Обучающие карты
- Интернет-ссылки
- Дополнительные задачи к отдельным главам,
- Таблицу случайных чисел
- Примеры и упражнения с использованием пакетов статистической обработки данных MicroCase и SPSS
- [illegible] диаграммы "Find the test"
- [illegible] переменных, измеряемых по порядковой шкале

Задачи

15.1 PS

Почему от выборов к выборам меняется процент проголосовавших избирателей? Для муниципальных выборов в пяти разных городах была собрана информация о проценте тех зарегистрированных избирателей, кто действительно проголосовал, а также об уровнях безработицы, образования (среднее количество лет, потраченных на учебу) и проценте негативной политической рекламы (т.е. рекламы, содержащей личные нападки на политических противников, изображение последних в негативном свете и т.п.). Для каждой зависимости требуется.

а. Построить диаграмму рассеяния и приблизительную (на глаз) прямую регрессии.

б. Рассчитать коэффициент наклона (b) и найти свободный член (a). *Совет: [illegible] вычисление b, затем перейти к вычислению a. При этом настоятельно рекомендуется все вычисления проводить в таблице, подобной таблице 15.3.*

в. Написать уравнение прямой регрессии, полученной по методу наименьших квадратов, и предсказать процент проголосовавших для города с уровнем безработицы 12%, города, в котором жители в среднем потратили на получение образования 11 лет, и для предвыборной кампании, в которой 90% агитации имели негативный по отношению к соперникам характер.

г. Рассчитать r и r^2. *Совет: настоятельно рекомендуется все вычисления проводить в таблице, подобной таблице 15.3. Если она использовалась для вычисления b, то у вас уже имеется большинство величин, необходимых для расчета r.*

[illegible] для каждой связи проверку на статистическую значимость.

е. В нескольких предложениях охарактеризовать тесноту и направление [illegible] дой связи. Какие из этих связей статистически значимы (если таковые [illegible] вообще имеются)? Какой фактор в наибольшей степени влияет на явку и [illegible]

Зависимость явки избирателей от уровня безработицы

	Явка избирателей	Уровень безработицы
A	55	5
B	60	8
C	65	9
D	68	9
E	70	0

Зависимость явки избирателей от уровня образования

	Явка избирателей	Среднее количество лет [illegible]
A	55	[illegible]
B	60	12 [illegible]
C	65	[illegible]
D	68	12.[illegible]
E	70	13.0

Зависимость явки избирателей от уровня [illegible] в избирательной [illegible]

Город	Явка избирателей	% негативной рекламы
[illegible]	[illegible]	48

15.2 SOC

В таблице для выборки из нескольких семей представлены данные о престиж[illegible]

Семья	Престижность работы отца	Престижность работы сына	Престижность работы дочери
[illegible]	[illegible]	[illegible]	8
E	69	75	[illegible]
F	[illegible]	72	[illegible]
G	64	60	[illegible]

Проанализировать связь между престижностью работы отца и сына [illegible] связь между престижностью работы отца и дочери. Для каждой связи требуется

а. Построить диаграмму рассеяния и приблизительную (на глаз) прямую регрессии.

б. Рассчитать коэффициент наклона (b) и найти свободный член (a).

в. Написать уравнение прямой регрессии, полученное по методу наименьших квадратов. Предсказать уровень престижности работы сына и дочери, у которых отец имел работу с уровнем престижности 72.

г. Рассчитать r и r^2.

д. Приняв, что данные семьи составляют случайную выборку, выполнить для обеих связей проверку на статистическую значимость.

е. В нескольких предложениях охарактеризовать тесноту и направление каждой связи. Оказывает ли престижность работы отца влияние на его детей? Одинаково ли это влияние для дочерей и сыновей?

15.3. G[illegible]

В ходе опроса жители дома для престарелых указали степень своей физической активности и количество посещений их родственниками и знакомыми за неделю. Существует ли зависимость между этими переменными для выборки из десяти респондентов, представленной в таблице? Построить диаграмму рассеяния и рассчитать r и r^2. Написать уравнение прямой регрессии, полученное по методу наименьших квадратов. Сколько посетителей в неделю может [illegible] дома престарелых с уровнем физической активности 5? А с уровнем 1[illegible]

Респондент	Уровень физической активности	Количество посетителей за неделю
A	[illegible]	[illegible]
B	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
D	[illegible]	[illegible]
E	15	[illegible]
F	9	7
G	7	10
H	[illegible]	15
I	[illegible]	[illegible]
J	[illegible]	[illegible]

15.4 PS

[illegible]

Избирательный участок	Процент демократов	Процент представителей национальных меньшинств	Явка избирателей
[illegible]	[illegible]	[illegible]	[illegible]
F	65	20	25
G	62	18	64
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

				Характеристики населения		
Штат	Убийства	Грабежи	Угоны автомобилей	Рост*	Плотность**	Урбанизация***
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Аризона	6	166	970	8.4	41	88
Вашингтон	4	120	568	4.7	85	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible] на период с 1990 по 1998 год

Численность населения на квадратную милю сухопутной территории штата по состоянию на 1998 год

[illegible] по состоянию на 1996 [illegible]

Источник: United States Bureau of the Census. Statistical Abstracts of the United States [illegible]

[illegible]

в. Написать уравнение [illegible] полученное по методу наименьших [illegible] уровня убийств следует ожидать в штате с [illegible]

роста населения [illegible] (т.е. убывание населения)? Каков уровень грабежей [illegible] от [illegible] плотностью населения 250? Предсказать уровень [illegible] автомобилей для штата, в котором 50% населения проживает в городах.

г. Рассчитать r и r^2.

д. [illegible] данные штаты составляют случайную выборку [illegible] связи и проверку на статистическую значимость.

е. В нескольких предложениях охарактеризовать силу и направление каждой связи.

15.6 PS

В [illegible] выборы городского совета. Для 15 случайным образом отобранных избирательных участков определить зависимость явки избирателей от процента представителей национальных меньшинств.

Избирательный участок	Процент проголосовавших от общего числа зарегистрированных избирателей	Процент представителей национальных меньшинств от общего числа зарегистрированных избирателей
A	[illegible]	0
B	47	15
C	65	75
D	[illegible]	10
E	[illegible]	5
F	64	25
G	50	97
H	49	45
[illegible]	45	7
[illegible]	4[illegible]	6
K	27	7
L	25	8
M	33	23
N	54	24
O	[illegible]	25

а. Рассчитать r и r^2.

б. Является ли r статистически значимым?

в. Охарактеризовать тесноту, направление и статистическую значимость связи.

15.7. SOC

Тренер баскетбольной команды небольшого колледжа считает, что его подопечные [illegible] играют и набирают больше очков, если выступают перед большим количеством зрителей. В таблице приведены количество набранных очков и посещаемость для всех матчей на своей площадке [illegible] сезон. Подтверждают ли эти данные мнение тренера?

[illegible]	Наблюдаемое [illegible]	Последовательность [illegible]
[illegible]	[illegible]	[illegible]

15.8. [illegible]

[illegible]	Расходы на образование (в расчете на одного человека)	Процент лиц с законченным средним образованием*	Рейтинг по доходу на одного человека
Арканзас	1008	77	46
[illegible]	[illegible]	[illegible]	[illegible]
Миссисипи	938	85	18
Небраска	[illegible]	77	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
Сев. Каролина	1013	64	7
Пенсильвания	1140	81	31
Вайоминг	[illegible]	64	[illegible]

[illegible]

15.9. [illegible] ду количеством детей и количеством часов, посвященных просмотру телевизора? [illegible] количеством детей? Между длительностью телепросмотров и престижностью работы? Какие из этих связей статистически значимы? Переменные и коды подробно поясняются в приложении Ж.

Престижность работы	Количество детей	Возраст	Частота посещения церкви	Длительность телепросмотров (часы)
12	3	34	3	1
50	0	41	0	3
7	0	52	7	2
69	3	67	0	5
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
52	0	22	2	3
[illegible]	[illegible]	[illegible]	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	5	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	5
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	3	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
46	[illegible]	[illegible]	[illegible]	[illegible]

Использование *SPSS for Windows* для вычисления r Пирсона

Демонстрационный пример с использованием SPSS 15.1. Какие факторы связаны с престижностью работы?

[illegible] *боты одного респондента)*

Если принадлежность человека [illegible] ет его собственную подготовку к у[illegible]

[illegible] другой стороны, если положение в обще[illegible] [illegible] в семье из какого социального класса он родился [illegible]

[illegible] Analyze [illegible] Correlate [illegible] Bivariate Correlations [illegible]

[illegible] Variables [illegible] Options [illegible]

[illegible] двусторонний критерий, проверку связей на статистическую значи[illegible] если пользователь предварительно не отключит эту опцию. Отметим, что ро Спирмена (см. главу 14) тоже является опцией. Щелкнуть по OK, при этом откры[illegible] окно вывода и на экран выводится следующая таблица.

Связи

		Количество полных лет обучения	Престижность работы отца (1980)	Престижность работы респондента (1980)
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible] ными, в том числе связь переменной с самой собой. Для [illegible] в верхней строке приведен[illegible] результаты проверки [illegible] значения, поскольку не все [illegible]

[illegible] трудностей, которые обусловлены социальным происхождением [illegible] papres80). [illegible] в семье с более высоким социальным положением, [illegible] может больше надеяться на успех в карьере, но тем не менее [illegible]

[illegible] более сложные [illegible] статистики, которые мы рассмотрим в главе 17.

Демонстрационный пример с использованием SPSS 15.2. Влияет ли пол человека на факторы, связанные с престижностью работы?

[illegible] ют? На [illegible] поиск ответа на этот вопрос [illegible] рассмот[illegible] в разделе "Демонстрационный пример с использованием SPSS 15.1" для мужчин и женщин. Если пол человека [illegible] на его профессиональный успех, то переменные *prestg80*, *educ*, *papres80* должны быть связаны между собой [illegible] для мужчин и женщин.

[illegible] социологическое исследование 1998 [illegible] В этом случае дан[illegible] по мужчинам и женщинам обрабатываются отдел[illegible] [illegible] в двух отдельных корреляционных матрицах — для мужчин и для женщин.

В главном меню щелкнуть по пункту **Data** (Данные), затем **Split File** (Разделить файл). В окне **Split File** (Разделить файл) щелкнуть по кнопке рядом с [illegible] by groups (Вывод организовать по группам). [illegible] для мужчин и женщин. В [illegible] списке переменных [illegible] по стрелке, перенести ее в окно **Groups Based On** (Группы, сформированные на основе признака). Щелкнуть по **OK**. Теперь любые [illegible] будут [illegible]

[illegible] требуется вновь вызвать окно **Split File** (Разделить файл) и щелкнуть по кнопке **Reset** (Восстановить). Теперь последовательно щелкнуть по **Statistics** (Анализ) [illegible] **Correlate** (Корреляции) [illegible] для переменных *educ*, *papres80*, *prestg80*. Открывается окно вы[illegible] и на экран выводится следующая таблица.

Для мужчин.

Связи

		Количество полных лет обучения	Престижность работы отца (1980)	Престижность работы респондента (1980)
[illegible]	[illegible]	[illegible]	[illegible]	0,46[illegible]
[illegible]	Количество наблюдений	0,302	000	0,[illegible]79
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

... женщин.		Связи		
		Количество полных лет обучения	Престижность работы отца (1980)	Престижность работы респондента (1980)
	Корреляция Пирсона	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	1,000	[illegible]
1980)	Уровень статистической значимости (двусторонний)	[illegible]	[illegible]	[illegible]
	Количество наблюдений	[illegible]	[illegible]	[illegible]
... 1980)	Уровень статистической значимости (двусторонний)	0,000	0,000	
	Количество наблюдений	710	566	

[illegible] умеренны и приблизительно одинаковы для обеих полов по силе. Правда, общественный класс женщин немного больше зависит от престижности отца (т.е. от того, к какому классу принадлежит семья, в которой выросла женщина) и образования, однако влияние указанных факторов настолько мало, что мы можем заключить: рассмотренные связи в незначительной степени зависят от пола человека.

Упражнения

15.1. Повторить пример 15.1 из раздела "Демонстрационный пример с использованием SPSS" или "Демонстрационный пример с использованием MicroCase", используя в качестве показателя положения в обществе переменную *income98* вместо *prestg80*. Использовать те же независимые переменные (*papres80*, [illegible]) и сравнить характеристики связей с теми, которые были определены в примере 15.1 (при этом можно не придавать значения тому обстоятельству, что переменная *income98* — порядковая.) В нескольких предложениях сформулируйте выводы.

15.2. Повторить пример 15.2 из раздела [illegible] ...нием SPSS" или "Демонстрацион[illegible] [illegible]

15.3. [illegible] С какими переменными [illegible] ...симые переменные (например, *age* и *educ*) [illegible] или четыре возможные н[illegible] [illegible]

Итоговые упражнения к части III

1. [illegible]

а. По [illegible] городам были собраны данные об общем уровне преступности (количество тяжких уголовных [illegible]) [illegible] ...ществует ли связь между этими переменными? [illegible]

[illegible]	[illegible]	0
[illegible]	1200	12
[illegible]	2000	9
[illegible]	1[illegible]00	11
[illegible]	1600	15
[illegible]	1000	20
[illegible]	[illegible]	[illegible]
[illegible]	1300	22
[illegible]	900	10
[illegible]	700	15

б. Имеется в наличии факты, что участие населения в общественной жизни [illegible] ...их районов (посещение и добровольные общественные организации, участие в местной политической жизни и т.д.) уменьшается на протяжении нескольких лет. Одной из причин этого явления называют телевидение. Подтверждают ли приведенные ниже данные такое мнение?

[illegible] участия в общественной жизни	Длительность телепросмотров		
	[illegible]	[illegible]	[illegible]
Низкая	5	10	16
Средняя	10	12	10
Высокая	15	8	[illegible]
Итого	30	30	3[illegible]

в. Один из общенациональных журналов составил рейтинг штатов [illegible] жизни (обычно) для оценки учитываемых уровень здравоохранения [illegible] учреждений для отдыха и развлечений, уровень безработицы и ряд других характеристик) и по уровню развития системы высшего образования. Оба [illegible] ...ли 20-балльные; где 1 ... соотв[illegible]

	Качество жизни	Уровень развития системы высшего образования
A	10	[illegible]
B	2	13
C	15	18
D	6	20
E	10	15
F	9	11
G	11	12
H	8	6
[illegible]	[illegible]	[illegible]
J	5	8

г. В Соединенных Штатах возрастает частота браков между представителями разных рас, в результате чего также возрастает количество людей смешанной расы. В таблице приведены данные о зависимости между комбинацией рас родите- [illegible] выборка. Существует ли статистическая связь между этими переменными?

[illegible]	Комбинация рас родителей			
	Чернокожий Белый	Белый Азиат	Чернокожий Азиат	Итого
[illegible]	2	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	6
[illegible]	[illegible]	[illegible]	4	[illegible]
[illegible]	20	15	10	45

7. [illegible] ких ситуаций и вопросов по ним. На каждый [illegible] ть с помощью, по крайней мере, одной из мето- [illegible] 15. Для каждой исследовательской ситуации тре- [illegible] наиболее подходящую меру связи и в нескольких предложе- [illegible] рая соответствующую меру связи, следует принимать во [illegible] возможных значений рассматриваемых переменных и их [illegible] ть к той или иной шкале.

Вопросы относятся к базе данных (см. таблицу), взятой из отчета Об[illegible]ненного [illegible] 17 [illegible] А. [illegible] В [illegible] были преобразованы (свернуты) в другие.

а. Существует ли связь между "частотой интимных отношений за последний [illegible] [illegible] переменные в данном примере — интервальные/относительные.

б. Зависит ли чувство страха респондента от его пола? От семейного положения?

в. [illegible]

Позиции, по которым проводился опрос

Семейное положение респондента (MARITAL)

1 Состоит в браке

Не состоит в браке (включая овдовевших, разведенных и т.д.)

[illegible] бые) Коды приведены в приложении Ж.

0 Никогда

1 Редко

2 Часто

Возраст респондента (AGE), (значения — фактическое количество лет)

Совокупный семейный доход респондента (INCOME98). Коды приведены в приложении Ж.

Пол

1 Мужской

2 Женский

Поддержка полового воспитания школьников (SEXEDUC)

1 За

2 Против

Боязнь ходить пешком ночью (FEAR)

Да

2 Нет

Частота интимных отношений за последний год. Коды приведены в приложении Ж.

Номер [illegible]	[illegible]	Посещение [illegible]	Возраст	Доход	Пол	Половое образование в школах	Страх	Частота интимных [illegible]

Часть IV

МНОГОМЕРНЫЕ МЕТОДЫ

16 Детальная разработка таблиц сопряженности

Цели изучения

К концу этой главы вы сможете

1. Объяснить сущность результатов многомерного анализа на основе наблюдения эффекта влияния управляющей переменной.
2. Выполнять построение и интерпретировать частные таблицы.
3. Вычислять и интерпретировать частные меры связи.
4. Распознавать и интерпретировать прямые, кажущиеся или опосредованные, а также интерактивные связи.
5. Вычислять и интерпретировать частный коэффициент гамма.
6. Перечислять и объяснять ограничения при детальной разра[illegible]

16.1 Введение

[illegible] ческий анализ, выполняемый только по двум переменным, может дать [illegible] на ограниченный круг вопросов, поэтому в типичном исследовательском [illegible] ируется большее количество переменных. В главах [illegible]2–15 мы выдел[illegible] [illegible] изучения парных связей могут использоваться различные статистические ме[illegible]. В этой главе, а также в главе [illegible]7, вы увидите, как можно расширить некоторые [illegible] лиза переменных, принадлежащих к номинальной или порядковой шкале, которые [illegible] вы познакомитесь с некоторыми м[illegible]тодами, которые могут использоваться для анализа переменных, принадлежащих к [illegible]

[illegible]

[illegible] о приходится собирать дополнительные сведения о какой-либо конкретной парной связи [illegible]

[illegible] чивых аргументов и окажутся очень полезными при проверке и переработке теории.

16.2 Управление с помощью третьей переменной

[illegible] данные представлены в виде таблицы, то в процессе выполнения многомерного анализа можно наблюдать за эффектами влияния других переменных на данную парную связь. Другими словами, мы можем наблюдать за связью между независимой X и зависимой Y переменными, контролируя третью переменную (которую мы будем обозначать Z). Для этого связь между переменными X и Y перестраивается с учетом каждого значения или категории переменной Z. Если управляющая переменная будет оказывать влияние, то связь между переменными X и Y будет изменяться в зависимости от состояния переменной Z.

Для наглядности предположим, что исследователю необходимо проанализировать связь между степенью интеграции индивидуума в группу или организацию и тем, насколько удовлетворен индивидуум данной группой. Исследователь решил сфокусироваться на студентах и степени их удовлетворенности учебой в колледже в целом. Необходимая информация о степени удовлетворенности Y была взята из выборки 270 студентов, которым был задан вопрос об их участии в деятельности студенческих клубов или организаций. Уровень интеграции X можно получить, разделив студентов на две категории: к первой категории будут относиться студенты, не являющиеся членами какой-либо организации (не члены), а ко второй категории будут относиться студенты, являющиеся членами хотя бы одной организации. [illegible] со студентами, не входящими ни в одну организацию. Информация о связи между этими двумя переменными приведена в таблице 16.1.

[illegible] ющиеся членами организаций, показывают высокую степень удовл[illegible] бой. Существование и направление связи [illegible]

[illegible]

[illegible] состоянием студента, причем эти результаты могут служить [illegible] для причинной или прямой связи между двумя переменными [illegible] следственная связь символически показана на рис. 16.1, где ст[illegible] влияние переменной X на переменную Y.

Таблица 16.1. Степень удовлетворенности учебой в зависимости от членства в студенческих организациях

	Членство в студенческих организациях (X)		
[illegible]ости учебой (Y)	Нет	Хотя бы в одной организации	Всего
[illegible]	48 (45,7%)	108 (65,5%)	[illegible]
[illegible]	105 (100%)	165 (100%)	[illegible]
	Гамма = 0,40		

X ⟶ Y

Рис. 16[illegible]

Разумеется, исследователь обнаружит, что членство не является единственным фактором, связанным со степенью удовлетворенности (если бы это было так, то [illegible]

переменные, поэтому исследователю предстоит систематически и логически анализировать влияние третьих переменных (Z-ых). В результате этого исследователь получит дополнительную информацию и подробные сведения о парной связи и её последствии [illegible] причинно-следственной связи между переменными X и Y.

[illegible]

денты, получающие более высокую степень [illegible] балл отметок, будут в боль[illegible]

[illegible] ким образом мож[illegible]

Давайте снова обратимся к та[illegible] представлена информация о рис[illegible]. Анализируя таблицу, можно ви[illegible]деть, что 165 студентов являются членами хотя бы одной организации, что 48 [illegible] дентов не являются членами ни одной организации [illegible] удовлетворенности учебой [illegible]

[illegible] правильном порядке вычисления коэффициента гамма. Для получения [illegible] данной мере связи см. главу 14.

[illegible] значимыми и показывающие высокую [illegible] есть высокий средний балл отметок [illegible]

[illegible]ельные наблюдения этой переменной не учитывается.

[illegible]

изменять свои значения. Для этого мы сортируем все наблюдения в выборке в соответствии с их значением по третьей переменной Z и изучаем связь между переменными X и Y для каждого значения (или долю) переменной Z. Для рассматриваемого примера мы построим отдельные таблицы, показывающие связь между членством в студенческой организации и степенью удовлетворенности учебой для каждой категории переменной среднего балла отметок. Такие таблицы называются **частными таблицами сопряженности**; вы их можете видеть в таблице 16.2. В этих таблицах студенты классифицированы по двум категориям — студенты, имеющие высокий и низкий средний балл отметок, причем в каждой категории представлены [illegible] 135 студентах. Чтобы получить две частные таблицы в рамках этих категорий, была выполнена повторная классификация студентов — по членству в студенческих организациях и по степени удовлетворенности учебой.

Таблица 16.2. Степень удовлетворенности учебой в зависимости от членства в студенческих организациях и с учетом контроля над средним баллом отметок

А. Высокий средний балл отметок

	Членство в студенческих организациях (X)		
Степень удовлетворенности учебой (Y)	Нет	Хотя бы в одной организации	Всего
[illegible]	[illegible]	[illegible]	[illegible]
	Гамма = 0,40		

Б. Низкий средний балл отметок

	Членство в студенческих организациях (X)		
Степень удовлетворенности учебой (Y)	Нет	Хотя бы в одной организации	Всего
[illegible]	[illegible]	[illegible]	[illegible]
	Гамма = 0,39		

[illegible] в данной ра[illegible]

[illegible] тальном, или уточненном виде. Частоты в клетках частных таблиц сопряженности [illegible]

[illegible] результат будет соответ-
вместе частоты в клетках частных [illegible] ствовать исходным частотам таблицы 16.1 [illegible] на то, как можно [illegible]

[illegible] ний управляющих переменных. Так, если бы мы контролировали пол студентов и [illegible]

[illegible] многомерный анализ начинается с построения частных таблиц сопряженности и [illegible] ной Z, в процессе которого выполняется сравнение частных таблиц друг с другом и [illegible] связь.

16.3. Интерпретация частных таблиц сопряженности

[illegible] приведены виды этих связей, определяемых путем сравнения частных таблиц сопря- [illegible]

1. Прямые [illegible] и Y остается неизменной во всех ча- [illegible]
2. Каж [illegible] связь между переменными X и Y оста- [illegible] сопряженности [illegible] чем в таблице со- [illegible]
3. [illegible]

Каждый тип связи имеет разный смысл для связей между переменными и в дальнейшем курсе статистического анализа. Сейчас каждый тип связи будет описан детально, а затем в таблице 16.5 будут подведены итоги.

Прямые связи

Такой характер связи часто называют репликацией, поскольку частные таблицы [illegible] ются неизменными как в частных таблицах, так и в таблице сопряженности. Мера связи, вычисленная для частных таблиц, имеет то же значение, что и мера связи, вычисленная для исходной таблицы сопряженности.

Такой результат показывает, что управляющая переменная Z не оказывает влияние на связь между переменными X и Y. Пример такого результата можно увидеть в таблице 16.2. В ней приводятся результаты изучения связи между членством в студенческих организациях X и степенью удовлетворенности учебой Y с учетом управления переменной Z (средний балл отметок). В таблице 16.2А показана связь для студентов, имеющих высокий средний балл отметок, а в таблице 16.2В показана связь для студентов, имеющих низкий средний балл отметок. Частные таблицы показывают одинаковые условные распределения Y для студентов с высоким и низким средним баллом отметок: приблизительно 45% студентов, не являющихся членами студенческих организаций, показывают высокую степень удовлетворенности учебой против приблизительно 66% студентов, являющихся членами студенческих организаций. Такой же характер связи наблюдался и в таблице сопряженности (табл. 16.1). Таким образом, условные распределения переменной Y остаются неизменными в каждой частной таблице и такими же, как и в исходной таблице сопряженности.

Характер частот, приводимых в клетках частных таблиц, будет проще определить, если рассчитать меру связи. Исходя из частот в клетках, представленных в таблице 16.2, значение коэффициента гамма для студентов, имеющих высокий средний балл отметок, составляет 0,40, а для студентов, имеющих низкий средний балл отметок, составляет 0,39. Значение парного коэффициента гамма (табл. 16.1) равно 0,40, и существенная эквивалентность значений гамма подтверждает, что связь между X и Y остается неизменной в частных таблицах и в таблице сопряженности.

[illegible] вает существенное влияние на парную связь: если бы это было так, то характер связи в частных таблицах отличалась бы от характера связи в таблице сопряженности. Поэтому эту переменную можно игнорировать при проведении любых последующих анализов. Опираясь на исходную задачу исследования, исследователь может прий- [illegible] ее всего выразят высокую степень удовлетворенности учебой в университете вне зависимости от среднего балла отметок. Размер денежных поощрений, выплачиваемых студентам (на основании их успеваемости), не влияет на связь между членством в студенческих организациях и степенью удовлетворенности учебой в университете. Студенты, имеющие низкий средний балл отметок и состоящие хотя бы в одной студенческой организации, выразят, вероятнее всего, такую же высокую степень удов-

[illegible] ниже высокий средний балл отметок и [illegible] табл. 16.1 и 16.2

[illegible]

[illegible] слабее, чем в исходной таблице сопряженности, однако во всех частных таб[illegible]

[illegible]

Для наглядности предположим, что в нашем примере исследователь [illegible]

[illegible] курса студенты [illegible] курса и на [illegible] курсы

[illegible]

[illegible] будут, вероятнее всего, [illegible] Особенно это касается тех случаев, когда деятельность организации связана с [illegible]

[illegible]

Практический пример 16.1

В выборке 75 юношей была [illegible]

[illegible]

двумя переменными для данной [illegible]

[illegible]

Степень причастности к совершению преступлений	Характеристика из средней школы		
[illegible]	[illegible]	[illegible]	[illegible]

Степень причастности к совершению преступлений	Характеристика из средней школы		
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

Степень причастности к совершению преступлений	Характеристика из средней школы		
[illegible]	[illegible]	[illegible]	[illegible]

Для юношей, проживающих в городских районах, связь между характеристикой со средней школы и степенью причастности к совершению преступлений не [illegible]. Значение коэффициента гамма для этой таблицы составляет –0,05, а проценты по столбцам являются очень схожими. Для данной группы степень причастности к совершению преступлений не связана с поведением и достижениями в средней школе.

[illegible] очень сильная связь между двумя переменными. Значение гамма составляет 0,89, при этом можно видеть резкое различие в процентах по столбцам. Плохие ученики, проживающие в сельской местности, более всего причастны к совершению преступлений.

Сравнение частных табл[illegible] позволяет обнаружить связь, [illegible] ря на то, что юноши, проживающие [illegible] совершению преступлений, чем юноши, проживающие в сельской [illegible] (7) 7% юношей, проживающих в городских районах, в высокой мере [illegible]

[illegible]

[illegible] к совершению преступлений никоим образом не связана с поведением и прилежанием в средней школе. Для юношей [illegible] характеристика со средней школы [illegible] ских районов более склонны к совершению преступлений, чем юноши из [illegible] [illegible]

Все рассуждения о возможных связях между этими переменными выражены в виде диаграммы, представленной на рис. 16.1. Отсутствие стрелки от переменной членства X к переменной удовлетворенности Y говорит о том, что они на самом деле не связаны между собой, а вызваны переменной старшинства курса, на котором учатся студенты Z. Если эта причинно-следственная диаграмма корректно описывает связь между тремя переменными (если связь между X и Y является кажущейся), тогда связь между членством в студенческих организациях и удовлетворенностью учебой должна исчезнуть после установления контроля над положением курса, на котором учатся студенты. То есть, даже если парный коэффициент гамма был бы равен 0,44 (таблица 16.1), все коэффициенты гамма, вычисленные в частных таблицах, будут иметь значения, близкие к нулю. В таблице 16.3 показаны частные таблицы сопряженности, построенные на основе управления над переменной уровня курса.

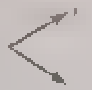


Рис. 16.2

[illegible]

[illegible] студенты младших курсов выразят, скорее всего, низкую степень удовлетворенности учебой [illegible] гамма, вычисленные по частным [illegible] и Y не имеют между собой [illegible] что переменные X [illegible] курса (переменная Z) является [illegible] удовлетворенности учебой, переориентировать и использовать переменную уровня курса [illegible]

Таблица 16.3. Степень удовлетворенности учебой в зависимости от членства в студенческих организациях и с учетом управления уровнем курса

А. Студенты старших курсов

Степень удовлетворенности учебой (Y)	Членство в студенческих организациях (X): Нет	Хотя бы в одной организации	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

B. Студенты младших курсов

Степень удовлетворенности учебой (Y)	Членство в студенческих организациях (X): Нет	Хотя бы в одной организации	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	2	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Гамма = 0,01

Результат (частные меры связи намного слабее исходной меры, однако равны [illegible]) [illegible] ассоциируется с другой идеей причинно-следственных связей между переменными. В дополнение к причинной схеме, в которой переменная Z рассматривается первичной для переменных X и Y, Z может находиться также и между двумя переменными. Такой характер связи называется также **интерпретацией**, а связь опосредованной. Схематически она показана на рис. 16.3, на нем видно, что переменная X имеет причинно-следственную связь с переменной Z, которая в свою очередь связана с переменной Y. Из этой схемы ясно, что, несмотря на наличие связи между переменными X и Y, связь эта осуществляется главным образом через управляющую переменную Z. Такой характер результатов не позволяет исследователю различать кажущиеся (рис. 16.2) и опосредованные связи (рис. 16.3). Дифференцировать эти два типа причинно-следственных связей можно лишь по времени либо теоретически, но не на основе количественной статистики (см. задачу *16.6б по кажущимся или опосредованным связям*).

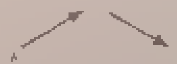


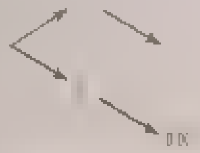
Рис. 16.3

Интерактивные связи

При таком характере связи, называемом также **спецификацией**, взаимосвязь между переменными X и Y существенно изменяется при различных значениях управляющей [illegible] переменными X и Y, чем [illegible] частной таблице связь между X и Y падает до нуля. Условно этот результат можно изоб-

разить так, как показано на рис 16.4, из которого видно, что переменная X и первая категория переменной Z (Z_1) оказывают сильное влияние на переменную Y, а вот во второй категории переменной Z (Z_2) связь между переменными X и Y отсутствует

Рис 16.4
Интерактивная связь

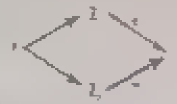


[illegible]

ную для борьбы с расовыми предубеждениями). Подобная программа повлияет, прежде всего, на белых служащих Z_1, чем на афроамериканцев Z_2. То есть, программа повлияет на расовые предубеждения только некоторых категорий субъектов. Частные таблицы сопряженности для белых служащих Z_1 могут показать сильную связь [illegible] служащих-афроамериканцев такая связь не будет показана. Служащие [illegible] в самом себе, будут в меньшей степени находиться под влиянием программы.

Интерактивная связь может принимать и другие формы. Например, связь между переменными X и Y может изменяться в частных таблицах не только в силе, но и по [illegible] Можно изобразить так, как показано на рис 16.5 — переменные X и Y имеют положительную связь с первой категорией переменной Z (Z_1) и отрицательную связь со второй категорией Z_2.

Рис 16.5
[illegible]



[illegible] что видно из процентного [illegible] студенческих организ[illegible] удовлетворенности учебой и меры связи (гамма [illegible] ых студентов составляет +0,67).

Таблица 16.4. Степень удовлетворенности учебой в зависимости от членства в студенческих организациях и с учетом контроля по расовой принадлежности

А. Белые студенты

	Членство в студенческих организациях (X)		
Степень удовлетворенности учебой (Y)	Нет	Хотя бы в одной организации	Всего
Низкая	40 (50,0%)	20 (16,7%)	60
Высокая	40 (50,0%)	100 (83,3%)	140

В. Темнокожие студенты

	Членство в студенческих организациях (X)		
Степень удовлетворенности учебой (Y)	Нет	Хотя бы в одной организации	Всего
Низкая	17 (68,0%)	36 (80,0%)	53
[illegible]	8 (32,0%)	9 (20,0%)	17
Всего	25 (100%)	45 (100%)	[illegible]
	Гамма = -0,31		

Для темнокожих студентов связь между членством в студенческих организациях и удовлетворенностью учебой в университете сильно отличается — эта модель практически противоположна модели для белых студентов. Подавляющее большинство темнокожих студентов (80%), являющихся членами студенческих организаций, показывает низкую удовлетворенность учебой, а значение коэффициента гамма для этой частной таблицы является отрицательным (-0,31). Так, для белых студентов степень удовлетворенности учебой возрастает при их членстве в студенческих организациях, тогда как для темнокожих студентов степень удовлетворенности учебой снижается, если они являются членами студенческих организаций. На основании этих результатов можно прийти к выводу, что социальные значения и последствия вступления студентов в студенческие клубы и организации будут различными в зависимости от расовой принадлежности, и что факт вступления в эти клубы и организации оказывает отличное влияние на темнокожих студентов. Может случиться, например, что темнокожие студенты вступают в организации, деятельность которых в корне отличается от организаций, в ряды которых входят белые студенты. Если более 1 [illegible] темнокожих студентов являются членами ассоциации, предубе[illegible] ассоциации может повысить степень неудовлетворенности учебой, поскольку в данном случае возрастает осознание межрасовых проблем (см. задачу 16.4 по интерактивной связи).

В заключение данного раздела следует отметить, что рассмотренные нами при[illegible] чем в [illegible] численные для частных таблиц вряд ли будут иметь такое [illegible]

для исходной таблицы (характер связи, при котором частные меры [illegible] [illegible] меньше по [illegible] [illegible] значением конкретной меры связи для всей таблицы, но не равны нулю, [illegible] является ослаблением). Вероятно, приводимые примеры лучше всего рассматривать в качестве идеальных примеров, с которыми впоследствии можно будет сравнивать любой эмпирический результат.

В таблице 16.3 приведены окончательные рекомендации по принятию решений относительно каждого результата, рассмотренного в этом разделе. Так как ваши собственные результаты будут, скорее всего, еще более неоднозначными, чем в рассмотренных примерах, вам следует отнестись к этой таблице как к набору советов, а не как к готовому руководству к действиям. Решая задачи, используйте собственные [illegible]

Таблица 16.3. Сводка возможных результатов управления с помощью третьей переменной

Частные таблицы по сравнению с исходной таблицей	Модель	Влияние на последующий анализ	Наиболее подходящий следующий шаг статистического анализа	Теория
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible] (категорий Z)	[illegible] теория того, что X является причиной Y, [illegible] учета переменной Z

16.4 Частный коэффициент гамма (G_p)

[illegible] силу связи [illegible] эффекта влияния управляющей переменной [illegible] частный гамма (G_p) [illegible] следующей формулы

Формула 16.1 [illegible]

где $\sum N_s$ — количество пар наблюдений, согласованно упорядоченных во всех частных таблицах сопряженности;

$\sum N_d$ — количество пар наблюдений, имеющих рассогласованно упорядоченных во всех частных таблицах сопряженности.

Другими словами, $\sum N_s$ является суммой всех N_s в частных таблицах, а $\sum N_d$ является суммой всех N_d в частных таблицах (схему вычисления N_s и N_d см. в гл. 14).

Чтобы показать схему вычисления G_p, давайте вернемся к таблице 16.2, в которой была показана связь между удовлетворенностью учебой и членством в студенческих организациях с учетом контроля над средним баллом отметок. Значения коэффициентов гамма, вычисленные в частных таблицах, были по существу равны между собой и равны значению парного коэффициента гамма для всей таблицы. Мы пришли к выводу, что управляющая переменная не оказывала влияние на связь, что было подтверждено также расчетом значения частного коэффициента гамма.

Из таблицы [illegible]	Из таблицы 16.2В
Высокий средний балл отметок	Низкий средний балл отметок
$N_s = (29)(54) = 1566$	$N_s = (28)(55) = 1540$
$N_d = (28)(24) = 612$	$N_d =$ [illegible]

$$= 1566 + 154[illegible] = 310[illegible]$$

$$[illegible]72 + 6[illegible]$$

$$G_p = \frac{\sum N_s - \sum N_d}{\sum N_s + \sum N_d}$$

$$[illegible] 344 [illegible]$$

$$[illegible] 1762 [illegible]$$

$$G_p = 0{,}4[illegible]$$

[illegible] организаций.

Когда мы управляли переменной старшинства курса (табл. 16.3), мы обнаружи- [illegible]

$N = (49)(12) = 588$

$N = (24)(24) = 576$

$\sum N = 776 + 588 = 1364$

$\sum N = 768 + 576 = 1344$

$G = 0.0$

16.5. Откуда берутся управляющие переменные?

Поэтому для того, чтобы исследовательский проект действительно был основан на некоторой теории, сама теория должна определять набор управляющих переменных, которые необходимо включить в анализ.

16.6. Ограничения при детальной разработке таблиц сопряженности

Основным ограничением данной методики является размер выборки. Для многомерного анализа детальная разработка является достаточно неэффективной методикой, поскольку исследователю приходится разбивать выборку на серию частных таблиц сопряженности. Если управляющая переменная имеет более двух-трех возможных значений или если мы попытаемся управлять несколькими [illegible] одновременно, для этого будут построены несколько частных таблиц сопряженности. С увеличением количества таких таблиц, имеющихся наблюдений окажется недостаточно, чтобы заполнить все клетки таблицы. В свою очередь, пустые клетки или недостаточное количество наблюдений в клетках могут стать причиной возникновения серьезных проблем при обобщении результатов и вызвать сомнения в достоверности полученных данных.

Пример, представленный в этой главе, начинался с двух дихотомических переменных и четырехклеточной исходной таблицы сопряженности (табл. 16.1). Каждая управляющая переменная также была дихотомической, и мы не производили сравнение более двух частных таблиц с четырьмя клетками в каждой или восемью клетками для каждой управляющей переменной. Если бы мы использовали управляющую переменную с тремя значениями, нам понадобилось бы заполнить 12 клеток, а если попытаться контролировать две дихотомические переменные, то в этом случае понадобятся 16 клеток в четырех различных частных таблицах. Понятно, что если управляющая переменная становится более детально разработанной и/или процесс контроля над переменными становится более сложным, то пустые клетки или [illegible] статочное их количество станут серьезной помехой.

Для разрешения этой дилеммы можно сразу же предложить два способа. Во-первых, можно просто сократить количество клеток в частных таб[illegible] категории переменных. Например, если все переменные будут дихотомическими, то количество клеток будет сведено к минимуму. [illegible]

[illegible] случайное решение не всегда является выполнимым (камнем преткновения могут оказаться [illegible]

[illegible]

[illegible] чем детальная разработка таблиц сопря[illegible] ваются все наблюдения одновременно [illegible]

16.7. Интерпретация статистик: анализ готовности принять участие в политических акциях

В главе 14 мы говорили об обязанностях граждан в США. В этом разделе мы рассмотрим вопрос, связанный с готовностью граждан принять участие в политических акциях. В качестве базы данных мы будем использовать базу данных всеобщего социального обследования за 1996 год (файл General Social Survey). Измерение зависимой переменной производилось по ответу на вопрос: Готовы ли Вы принять участие в митинге представителей общественности? Эта переменная имеет четыре категории отклика (ответа): определенно нет, скорее всего, нет; скорее всего, да, определенно да. Чтобы свести количество клеток к минимуму, эти категории были объединены в простую дихотомию: да и нет. В качестве независимой переменной используется общественный класс, а двумерная связь представлена в табл. 16.6.

Таблица 16.6. Готовность принять участие в политических акциях в зависимости от принадлежности к определенному общественному классу

Готовы ли вы принять участие в митинге представителей общественности?	Общественный класс				
	Низший	Рабочий	Средний	Высший	Всего
Нет	28 (39.4%)	235 (40.9%)	196 (34.3%)	18 (31.6%)	477 (37.4%)
Да	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Всего	71 (100.0%)	574 (100.0%)	572 (100.0%)	57 (100.0%)	1274 (100.0%)

Гамма = 0,13

Судя по значению парного коэффициента гамма, равного 0,13, можно предположить о существовании положительной, но слабой связи между принадлежностью к определенному общественному классу и готовностью принять участие в полити[illegible] 60,6% [illegible] рабочему [illegible] классу (59,1%), приблизительно одинаково ответили о готовности участвовать в по[illegible] 65,7% [illegible] (68,4%) [illegible]

[illegible]

[illegible] является умеренной [illegible] отсутствие существенной связи между готовностью для женщин (часть В) показывает [illegible] классом. Проценты по столбцам [illegible]

[illegible] для каждого класса, а значение гамма составляет –0,01. Сравнение [illegible] между собой и с исходной таблицей сопряженности показывает [illegible] [illegible] постепенно возрастает с переходом к более высокому общественному классу для мужчин и остается [illegible] для женщин.

Таблица 16.7 Готовность принять участие в политических акциях в зависимости от принадлежности к определенному общественному классу, с учетом контроля по полу респондентов (в процентном соотношении)

А. Мужчины

Готовы ли вы принять участие в митинге представителей общественности?	Общественный класс: Низший	Рабочий	Средний	Высший
Нет	[illegible]	[illegible]	[illegible]	[illegible]
Да	50,0%	67,7%	7_,7%	[illegible]
Всего	100,0% [illegible]	100,0% [illegible]	100,0% [illegible]	[illegible]

Гамма = 0,3[illegible]

Б. Женщины

Готовы ли вы принять участие в митинге представителей общественности?	Общественный класс: Низший	Рабочий	Средний	Высший
Нет	34,0%	39,7%	38,8%	40,3%
Да	64,0%	60,3%	61,2%	60,0%
Всего	100,0% (47)	100,0% (302)	100,0% (325)	100,0% (30)

Гамма = –0,1

В таблице 16.8 приведены результаты контроля [illegible] [illegible] респондентов связь также является положительной, но слабой (G = 0,07).

[illegible] между готовностью [illegible] результаты не [illegible] меньшинств в нашем обществе — [illegible] на готовность принять участие в [illegible]

Таблица 16.8 Готовность принять участие в митинге представителей общественности в зависимости от общественного класса и с учетом контроля по расовой принадлежности (в процентном соотношении)

А. Белые

Готовы ли вы принять участие в митинге представителей общественности?	Общественный класс: Низший	Рабочий	Средний	Высший
Нет	[illegible]	[illegible]	[illegible]	[illegible]
Да	6_,6%	57,5%	65,8%	67,3%
Всего	[illegible]	[illegible]	[illegible]	[illegible]

Гамма = 0,14

В. Небелые

Готовы ли вы принять участие в митинге представителей общественности?	Общественный класс: Низший	Рабочий	Средний	Высший
Нет	[illegible]	[illegible]	[illegible]	[illegible]
Да	[illegible]	[illegible]	[illegible]	[illegible]
Всего	100,0% (16)	100,0% (_19)	100,0% (88)	100,0% (5

Гамма = 0,07

В исходной версии было три категории: белые, темнокожие и другие. В связи с тем, что в последнюю категорию входили только несколько респондентов, автор объединил категории темнокожих [illegible]

Заключение

1. Многие вопросы, изучаемые в исследовании, требуют выполнения анализа взаимосвязей между множеством переменных, даже если исследователь сконцентрирован в первую очередь на конкретной парной связи. Многомерные статистические методы предлагают исследователю набор инструментов, посредством которых можно собрать дополнительные сведения об интересуемых переменных и благодаря которым можно изучить причинные взаимоотношения.
2. [illegible] сопряженности. Сравнение будет выполнять легче, если для [illegible]
3. Между независимой переменной X и зависимой переменной Y существует при[illegible] такой же, как и в исходной сопряженности [illegible] существовании причинно-следственной связи между переменными X и Y.
4. Если связь между переменными X и Y остается неизменной во всех частных таб[illegible]

6. Частичный коэффициент связи G_p является [illegible] статистикой, [illegible]

[illegible] необходимо установить контроль.

8. Контроль над третьими переменными посредством создания частичных таблиц [illegible] пытается контролировать несколько [illegible] можно решить, если совместить категории или осуществить сбор [illegible]

Сводка формул:

[illegible] 16.1 $G_p = \frac{\sum N_s - \sum N_d}{[illegible]}$

Словарь

Частичная разработка [illegible] и более детально проводятся или уточ-

Интерактивная связь. [illegible] связь, при которой двумерная связь [illegible] переменной

Частичная связь. См. опосредованная связь

Кажущаяся связь. [illegible] после установления контроля над третьей [illegible]

Объяснение. См. кажущаяся связь

Опосредованная связь. [illegible] щей переменной

Прямая связь. [illegible]

Репликация. См. прямая связь

Спецификация. См. интерактивная связь

Частичный коэффициент связи [illegible]

Частичная таблица сопряженности [illegible]

Z. Символ управляющей переменной

Мультимедийные ресурсы

[illegible] (Виртуальное сообщество Центра социологических ресурсов Ziegler)

[illegible]

- Обзор основ математики
- [illegible]
- Обучающие карты.
- Internet-ссылки.
- Дополнительные материалы к главам.
- Таблица случайных чисел.
- [illegible] SPSS
- Функциональные диаграммы Find the test
- Проверка гипотез для переменных [illegible]

Задачи

16.1. SOC

В главе 14 [illegible] рассмотрен [illegible] вопрос [illegible] небольшой выборки иммигрантов. В таблице были приведены данные о связи между продолжительностью пребывания в США и уровнем владения английским языком.

[illegible]	Продолжительность пребывания		
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	20	10	
Высокий	5	15	
Всего	25	20	

а. Если это необходимо, определите проценты по столбцам и вычислите для данной таблицы значение коэффициента гамма. Опишите парную связь на основании данных о ее тесноте и направлении.

Влияет ли пол иммигранта на связь между продолжительность[illegible] и уровнем владения языком? В следующих двух частных таблицах сопряженности показана парная связь для мужчин и женщин.

А. Мужчины

[illegible]	Продолжительность пребывания		
[illegible]	Менее пяти лет	Более пяти лет	Всего
Низкий	10	5	15
Высокий	2	8	10
Всего	[illegible]	[illegible]	[illegible]

Б. Женщины

Уровень владения английским языком	Продолжительность пребывания		
	Менее пяти лет	Более пяти лет	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	7	[illegible]
Всего	13	12	25

б. Определите проценты [illegible] [illegible] ости. Сравните проценты и [illegible] частных таблицах [illegible] ной таблицы сопряженности. Влия[illegible] парную связь контроль над по[illegible]

[illegible]

А. Выходцы из Азии

Уровень владения английским языком	Продолжительность пребывания		
	Менее пяти лет	Более пяти лет	Всего
Низкий	8	4	12
Высокий	2	5	7
Всего	10	9	19

Б. Выходцы из Латинской Америки

Уровень владения английским языком	Продолжительность пребывания		
	Менее пяти лет	Более пяти лет	Всего
Низкий	12	6	18
[illegible]	3	10	13
[illegible]	15	16	31

в. Определите проценты по столбцам и вычислите значения гамма для каждой частной таблицы сопряженности. Сравните проценты и значения гамма для частных таблиц сопряженности друг с другом и со значением гамма для исходной таблицы сопряженности. Влияет ли на парную связь контроль над переменной *происхождение иммигрантов*?

16.2. SOC

Во время проведения переписи населения в 100 районах были собраны данные о процентах случаев совершения самоубийств, возрастном составе и безработице. Переменные *процент случаев совершения самоубийств* и *процент безработицы* были преобразованы в дихотомические, используя в качестве точки разделения медиану, поэтому для каждого района переписи процент можно оценить как высокий или низкий. Возрастной состав измерялся в процентах населения в возрасте 65 лет и старше. Эта переменная также являлась дихотомической, поэтому процент для каждого района переписи также можно оценить как высокий или низкий. В следующих таблицах приведены данные о парной связи между процентом случаев совершения самоубийств и возрастном составе населения, и данные о такой же связи, но с учетом контроля над безработицей.

Процент случаев совершения самоубийств	Процент населения в возрасте 65 лет и старше		
	Низкий	Высокий	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

а. Определите проценты и значение гамма для таблицы сопряженности. Опишите парную связь на основании данных о тесноте и направлении.

[illegible] процент с[illegible] совершения самоубийств [illegible] и с учетом контрольной [illegible] безработицей

А. Высокий уровень безработицы

Процент случаев совершения самоубийств	Процент населения в возрасте 65 лет и старше: Низкий	Высокий	[illegible]
[illegible]	23	10	[illegible]
Высокий	5	2	[illegible]
Всего	28	22	[illegible]

Б. Низкий уровень безработицы

Процент случаев совершения самоубийств	Процент населения в возрасте 65 лет и старше: Низкий	Высокий	[illegible]
Низкий	22	10	32
Высокий	5	13	[illegible]
Всего	27	23	[illegible]

б. Определите проценты и значения коэффициента гамма для каждой частной таблицы сопряженности. Сравните частные таблицы друг с другом и с [illegible] коэффициента гамма (G_p).

в. Опишите результаты управления переменной. Влияет ли каким-либо образом безработица на связь между возрастным составом населения [illegible] с управляемой пере[illegible] возможные причинно-следственные связи между этими переменными.

16.3. SW

[illegible]

Влияет ли на эту связь пол пациентов?

Степень восприятия дейст[illegible]

[illegible]

[illegible] этом контроле по полу пациентов

А. Ж[illegible]

Степень восприятия действительности	Продолжительность курса [illegible]: Менее 5 [illegible]	Более 5 [illegible]	Всего
[illegible]	[illegible]	[illegible]	[illegible]
Высокая	60	37	97
Всего	155	157	312

Б. Мужчины

Степень восприятия действительности	Продолжительность курса [illegible]: Менее 5 лет	Более 5 [illegible]	Всего
Низкая	105	93	198
П[illegible]	[illegible]	[illegible]	[illegible]
Всего	162	134	296

16.4. SOC

Существует ли связь между возрастом и отношением к сексу? Действительно ли люди пожилого возраста более консервативно относятся к вопросам, относящимся к половой жизни? 925 респондентов попросили высказать свое мнение о добрачной половой жизни. Полученные ответы были объединены в две категории: для тех респондентов, которые полагают, что добрачная половая жизнь это всегда неправильно и для тех, кто считает, что это бывает нормально в силу некоторых обстоятельств (иногда неправильно). Результаты опроса представлены в сводной таблице с разбивкой по возрасту.

Добрачная половая жизнь это:	Возраст респондентов: Младше 35 лет	35 лет и старше	Всего
Всегда неправильно	[illegible]	[illegible]	[illegible]
Н[illegible]	[illegible]	[illegible]	[illegible]
В[illegible]	[illegible]	[illegible]	[illegible]

а. Определите проценты и значение коэффициента гамма для таблицы. Существует ли связь между этими двумя переменными? Опишите парную связь на основании данных о силе и направлении. Далее показана парная [illegible] после установления контроля над полом респондентов. Оказывает [illegible] респондентов?

А. Мужчины

Добрачная половая жизнь это:	Возраст респондентов: Младше 35 лет	35 лет и старше	Всего
В[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]	Возраст респондентов		
	Молодые [illegible]	35 лет и старше	Всего
[illegible]	20	[illegible]	[illegible]
[illegible]	[illegible]	100	[illegible]
[illegible]	[illegible]	20	[illegible]

[illegible]

16.5. [illegible]

[illegible] года: (1) закончил ли обучаемый учебную программу, (2) [illegible] устроиться на работу и проработать на ней в течение года после окончания курсов, (3) пол обучаемого, (4) расовая принадлежность обучаемого. Связано ли трудоустройство с окончанием учебных курсов? Влияет ли на связь между окончанием учебных курсов и трудоустройством расовая принадлежность или пол обучаемых?

Трудоустройство в зависимости от обучения

Смогли ли Вы устроиться на работу и проработать на ней в течение одного года после окончания курсов?	Завершено ли обучение?		
	Да	Нет	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Трудоустройство в зависимости от обучения и с учетом контроля над расовой принадлежностью [illegible]

А. Белые

Смогли ли Вы устроиться на работу и проработать на ней в течение одного года после окончания курсов?	Завершено ли обучение?		
	Да	Нет	Всего
[illegible]	65	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Б. [illegible]

Смогли ли Вы устроиться на работу и проработать на ней в течение одного года после окончания курсов?	Завершено ли обучение?		
	Да	Нет	Всего
[illegible]	6	27	[illegible]
[illegible]	[illegible]	79	[illegible]
[illegible]	64	106	[illegible]

Трудоустройство в зависимости от обучения и с учетом контроля над полом респондентов

А. [illegible]

Смогли ли Вы устроиться на работу и проработать на ней в течение одного года после окончания курсов?	Завершено ли обучение?		
	Да	Нет	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Б. [illegible]

Смогли ли Вы устроиться на работу и проработать на ней в течение одного года после окончания курсов?	Завершено ли обучение?		
	Да	Нет	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

16.6. [illegible]

Что движет людьми, вступающими в ряды защитников окружающей среды? В ходе недавно проведенного исследования была получена информация о степени заинтересованности в вопросах глобального потепления и выпадения кислотных дождей. Связана ли заинтересованность в вопросах защиты окружающей среды с уровнем образования респондентов? Какое влияние оказывает управляющая переменная? Напишите свои заключения.

Степень заинтересованности в вопросах защиты окружающей среды в зависимости от уровня образования

Степень заинтересованности в вопросах защиты окружающей среды	Уровень образования		
	Низкий	Высокий	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
Всего	49	83	132

а. Степень заинтересованности в вопросах защиты окружающей среды в зависимости от уровня образования и с учетом контроля над полом респондентов

А. [illegible]

Степень заинтересованности в вопросах защиты окружающей среды	Уровень образования		
	Низкий	Высокий	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

Степень заинтересованности в вопросах защиты окружающей среды	Уровень образования: Низкий	Высокий	Всего
[illegible]	13	18	31
Низкая	11	26	37
Всего	[illegible]	44	68

[illegible] ...енности в вопросах защиты окружающей среды в за-... [illegible]

A. Низкий уровень доверия:

Степень заинтересованности в вопросах защиты окружающей среды	Уровень образования: Низкий	Высокий	Всего
[illegible]	6	22	28
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	6	[illegible]	[illegible]

Б. [illegible]

Степень заинтересованности в вопросах защиты окружающей среды	Уровень образования: Низкий	Высокий	Всего
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	2	8	[illegible]
[illegible]	33	[illegible]	[illegible]

[illegible]

A. Белые

Степень заинтересованности в вопросах защиты окружающей среды	Уровень образования: Низкий	Высокий	Всего
[illegible]	9	[illegible]	[illegible]
[illegible]	18	[illegible]	[illegible]
[illegible]	3	[illegible]	[illegible]

Б. [illegible]

Степень заинтересованности в вопросах защиты окружающей среды	Уровень образования: Низкий	Высокий	Всего
[illegible]	9	[illegible]	[illegible]
[illegible]	4	4	[illegible]
[illegible]	2	[illegible]	[illegible]

6.7. SOC

[illegible]

...ная переменная пол на парные связи. В приводимой далее таблице сопряженности и в частных таблицах сопряженности указаны значения парных коэффициентов гамма, вычисленных в предыдущем примере. Вычислите проценты и значения показателя гамма для каждой частной таблицы и сформулируйте [illegible] Оказывает [illegible] ...пондента? Таблица сопряженности (из задачи 14.15).

Поддерживаете ли Вы легализацию абортов?	Уровень годового дохода: Менее $24900	От $24900 до $50000	Более $50000	Всего
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Нет	365	299	250	9[illegible]5
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

5. [illegible]

A. [illegible]

Поддерживаете ли Вы легализацию абортов?	Уровень дохода: Менее $24900	От $24900 до $50000	Более $50000	Всего
Да	89	96	[illegible]	[illegible]
Нет	130	141	122	393
[illegible]	[illegible]	23[illegible]	[illegible]	[illegible]

Б. [illegible]

Поддерживаете ли Вы легализацию абортов?	Уровень дохода: Менее $24900	От $24900 до $50000	Более $50000	Всего
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible] (из задачи 14.15)

Поддерживаете ли Вы легализацию абортов?	Уровень дохода: Менее $24900	От $24900 до $50000	Более $50000	Всего
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

	Уровень дохода			
Придерживаетесь ли Вы ...	Менее $24900	От $24900 до $50000	Более $50000	Всего

	Уровень дохода			
Придерживаетесь ли Вы ...	Менее $24900	От $24900 до $50000	Более $50000	Всего

...диционным распределением ролей в семье

- Опишите результаты анализа.

	Уровень дохода			
Согласны ли Вы с традиционным распределением ролей в семье между мужчиной и женщиной?	Менее $22500 1	От $22500 до $50000 2	Более $50000 3	Всего

	Уровень дохода			
Согласны ли Вы с традиционным распределением ролей в семье между мужчиной и женщиной?	Менее $22500 1	От $22500 до $50000 2	Более $50000 3	Всего

	Уровень дохода			
Согласны ли Вы с традиционным распределением ролей в семье между мужчиной и женщиной?	Менее $22500 1	От $22500 до $50000 2	Более $50000 3	Всего
Всего	354	299	265	918

	Уровень дохода			
Согласны ли Вы с традиционным распределением ролей в семье между мужчиной и женщиной?	Менее $22500 1	От $22500 до $50000 2	Более $50000 3	Всего
Всего	26			48

...соответствующими значениями. Не все респонденты ответили на каждый заданный вопрос.

Согласны ли Вы с традиционным распределением ролей в семье между мужчиной и женщиной?	Уровень дохода			
	Менее $22500 (1)	От $22500 до $50000 (2)	Более $50000 (3)	Всего
[illegible]	[illegible]	66	8	[illegible]
[illegible]	4	73	[illegible]	[illegible]
[illegible]	[illegible]	43	[illegible]	[illegible]

Согласны ли Вы с традиционным распределением ролей в семье между мужчиной и женщиной?	Уровень дохода			
	Менее $22500 (1)	От $22500 до $50000 (2)	Более $50000 (3)	Всего
[illegible]	2	64	78	4[illegible]
[illegible]	9	9	149	4[illegible]
[illegible]	[illegible]	79	[illegible]	[illegible]

Мнение о традиционном распределении ролей в семье между мужчиной и женщиной в зависимости от уровня дохода и с учетом контроля над возрастным [illegible]

Согласны ли Вы с традиционным распределением ролей в семье между мужчиной и женщиной?	Уровень дохода			
	Менее $22500 (1)	От $22500 до $50000 (2)	Более $50000 (3)	Всего
[illegible]	36	23	6	[illegible]
[illegible]	26	68	48	[illegible]
[illegible]	61	5	54	30[illegible]

Согласны ли Вы с традиционным распределением ролей в семье между мужчиной и женщиной?	Уровень дохода			
	Менее $22500 (1)	От $22500 до $50000 (2)	Более $50000 (3)	Всего
[illegible]	62	[illegible]	4	[illegible]
[illegible]	19	208	2[illegible]	[illegible]
[illegible]	195	[illegible]	285	[illegible]

Согласны ли Вы с традиционным распределением ролей в семье между мужчиной и женщиной?	Уровень дохода			
	Менее $22500 (1)	От $22500 до $50000 (2)	Более $50000 (3)	Всего
[illegible]	[illegible]	94	5	[illegible]
[illegible]	[illegible]	62	99	2[illegible]
[illegible]	222	186	156	5[illegible]

Использование *SPSS for Windows* для детальной разработки таблиц сопряженности

Демонстрационный пример с использованием SPSS 16.1. Анализ влияния возраста и пола респондентов на отношение к добрачной половой жизни

[illegible] Crosstabs [illegible] Crosstabs [illegible] Crosstabs [illegible] Crosstabs [illegible] сти экрана.

В демонстрационном примере с использованием SPSS 12.3 и 14.1 мы изучали связи между переменной *premarsx* и перекодированной переменной *age* или *agecat* [illegible]

[illegible] Crosstabs [illegible] Проверьте, включена ли опция вычисления значений гамма и процентов по столб[illegible]

[illegible] что мужчины относятся к добрачной половой жизни лояльнее по сравнению с жен[illegible]

Демонстрационный пример с использованием SPSS 16.2

Действительно ли престижная работа приносит более высокую степень удовлетворенности? Другой взгляд на вопрос

[illegible] SPSS [illegible]

[illegible] Введем некоторые дополнительные управляющие переменные, например пол [illegible]

[illegible] изучим первую связь, однако теперь в качестве управляющих переменных бу- [illegible]

[illegible] удовлетворенности работой с фактором престижа изменяется в зависимости от пола респондентов? Действительно ли уровень образования респондента влияет на связь между степенью удовлетворенности работой и фактором престижа?

Переменная *degree* имеет в общей сложности пять значений — слишком много, чтобы эту переменную можно было использовать в качестве управляющей переменной. Чтобы упростить переменную, все респонденты, получившие какую-либо ученую степень в колледже, были объединены в одну категорию. Схема перекодировки выглядела следующим образом:

0 → 0 (уровень образования меньше средней школы)

1 → 1 (имеющие аттестат об окончании средней школы)

2 thru 4 → 2 (образование на уровне колледжа)

Перекодированной версии этой переменной было присвоено имя *rdeg*. В демонстрационном примере 14.3 приведены схемы перекодирования переменных *[illegible]* [illegible]

Запустите процедуру **Crosstabs** (Таблицы сопряженности) и в окне **Row Variable** [illegible] **Column Variable** [illegible] ные таблицы сопряженности. Вместо этого сконцентрируемся на значениях гамма.

	Гамма
Для таблицы сопряженности	
[illegible]	
[illegible] уровнем ниже средней школы	-0.29
Для людей, имеющих образование на уровне колледжа	0.26

[illegible] связь. Степень удовл- [illegible] уровней образования [illegible] слабой. Судя по [illegible]

процентов по столбцам, фактор престижа оказывает меньшее влияние на степень удовлетворенности работой для менее образованных респондентов.

Контроль над переменной пол также показывает некоторое взаимодействие.

	Гамма
[illegible]	
Мужчины	-0.33
Женщины	0.20

Как для мужчин, так и для женщин очень престижная работа связана с более высокой степенью удовлетворенности, однако это влияние сильнее проявляется для мужчин. Проценты по столбцам также показывают, что фактор престижа работы по-разному влияет на степень удовлетворенности работой для мужчин и для женщин.

Упражнения

16.1 Используя демонстрационный пример SPSS 16.1, проанализируйте связь между перекодированными переменными *age*, *cappun*, *grass* (отношение к легализации слабых наркотиков) и *gunlaw* (отношение к свободному ношению оружия), используя переменную *sex* в качестве управляющей переменной. Опишите полученные результаты. Убедитесь в том, что связи характеризуются как прямые, кажущиеся или интерактивные.

16.2. Используя демонстрационный пример SPSS 16.2, проанализируйте связь между уровнем дохода (*income91*) и степенью удовлетворенности работой, контролируя переменную *sex* и преобразованную переменную *degree* (ученая степень). Преобразуйте переменную *income91* в дихотомическую перед ее включением в анализ, используя в качестве точки разделения медиану. Опишите полученные результаты. Убедитесь в том, что связи характеризуются как прямые, кажущиеся или интерактивные.

17 Частная корреляция, множественная регрессия и корреляция

Цели изучения

Изучив настоящую главу, вы сможете

1. Вычислять и интерпретировать коэффициенты частной корреляции.
2. Получать, используя метод наименьших квадратов, уравнение множественной регрессии, а также интерпретировать как само уравнение, так и частные коэффициенты регрессии.
3. Вычислять и интерпретировать коэффициент множественной корреляции (R).
4. Объяснять ограничения в анализе частной и множественной регрессии.

17.1. Введение

В главе 16 было сказано, что статистический анализ, выполняемый с использованием только двух переменных, может дать ответ лишь на ограниченный круг вопросов. По своей природе социологическое [illegible] [illegible] [illegible] [illegible] используемые при анализе причинно-следственных связей [illegible] предсказаний — наиболее [illegible] [illegible] [illegible] [illegible] ют самые разнообразные способы, позволяющие [illegible] объем информации и предлагающие [illegible]

Вначале мы рассмотрим анализ [illegible] тоду управления с помощью третьей [illegible] посредством построения частных таблиц сопряженности (см. главу 16). Еще один метод основан на множественной регрессии и корреляции и позволяет исследователю оценить эффекты влияния — по отдельности, так и в комбинации — одной или нескольких [illegible] менных на зависимую переменную.

В настоящей главе мы будем рассматривать ситуации, в которых анализ выполняется только по трем переменным. Это наименее сложный вариант применения данных методов, хотя расширить эти методы на ситуации анализа четырех и более переменных можно сравнительно просто. Чтобы избежать трудностей при выполнении вычислений, необходимых в более сложных ситуациях, я советую вам воспользоваться любым компьютерным пакетом статистической обработки данных, который наверняка имеется в вашем вычислительном центре.

17.2. Частная корреляция

Метод частной корреляции может быть использован в тех случаях, когда исследователю необходимо проследить поведение определенной парной связи в присутствии третьей переменной. Наблюдая за коэффициентами частной корреляции, можно выявить прямые, кажущиеся или опосредованные связи (см. главу 16 и табл. 16.5), а [illegible] ми. Однако в отличие от методов, рассмотренных в главе 16, при анализе частной корреляции исключена возможность проверки парной связи по каждой категории управляющей переменной. По этой причине методы частной корреляции не позволяют обнаружить интерактивные связи между переменными — связи, в которых характер парной связи различен в разных категориях управляющей переменной.

Прежде чем перейти к обсуждению метода расчета частной корреляции, следует познакомиться с некоторыми новыми терминами. В связи с тем что мы будем работать с несколькими парными связями, для различения корреляций мы будем использовать подстрочные индексы. Так, символ r_{yx} будет обозначать коэффициент корреляции между переменными Y и X. Подобным образом коэффициент корреляции между переменными Y и Z будет иметь обозначение r_{yz}, а коэффициент корреляции между переменными X и Z будет иметь обозначение r_{xz}. Корреляции, вычисленные для парных связей, часто называются **корреляциями нулевого порядка**.

Коэффициенты частной корреляции называются коэффициентами частной корреляции первого порядка, если осуществляется контроль над одной переменной, они имеют обозначение $r_{yx.z}$. Переменная, указанная справа от точки, называется управляющей переменной. [illegible] корреляции, измеряющему связь между переменными X и Y при условии контроля над переменной Z. Формула для вычисления коэффициента частной корреляции первого порядка [illegible]

Формула 17.1 [illegible]

Обратите внимание, прежде чем производить расчеты по этой формуле, сначала необходимо вычислить коэффициенты нулевого порядка между всеми возможными парами переменных (между переменными X и Y, X и Z, Y и Z).

Чтобы понять порядок вычисления коэффициентов частной корреляции первого порядка, давайте вернемся к задаче из главы 15, в которой мы проверяли связь

[illegible] объемом работ по дому Y, выполняе[illegible] [illegible] которых оба работают [illegible] порядка между этими двумя переменными ($r_{yx} = 0,50$) [illegible] положительной связи. Допустим, исследовате[illegible] [illegible] влияние, оказываемые на парную корреляцию [illegible] экономический статус (SES). Исход[illegible] [illegible]

Таблица 17.1. Значения трех переменных для 12 семей, в которых оба родителя имеют работу, и корреляции нулевого порядка

Семья	Объем работ по дому, выполняемой главой семейства, Y	Количество детей в семье, X	Количество лет, потраченных главой семейства на обучение, Z
[illegible]	[illegible]	[illegible]	12
В	[illegible]	[illegible]	14
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	2	[illegible]
Г	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	12
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	4	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	0	[illegible]	[illegible]

[illegible]

$$r_{yx.z} = \frac{[illegible]}{\sqrt{1-0,09}\sqrt{1-0,22}}$$

$$= \frac{0,36}{[illegible]}$$

[illegible]

Коэф[illegible] [illegible]чению, чем коэффициент нулевого порядка ($r_{yx} = 0,50$), однако разница между ними не очень велика. Этот результат позволяет сделать предположение о существовании [illegible] [illegible] с увеличением количества детей в семье. Следующим этапом статистического ана[illegible] [illegible] переменной (см. табл. 16.5). Чем дольше парная связь будет сохранять свою силу в серии контроля над третьими переменными (Z переменными), тем очевиднее существование прямой связи между переменными X и Y.

[illegible] менной Z, поэтому следующим этапом статистического анализа будет включение управляющей переменной в анализ в качестве независимой переменной (см. табл. 16.5). Необходимо отметить, что вычисление коэффициентов частной корреляции не может различить случаи, в которых управляющая переменная является первичной (Z порождает X и Y), и случаи, в которых управляющая переменная является промежуточной переменной (X порождает Z, которая в свою очередь порождает Y). Делать вывод о возможных причинных связях следует лишь на основании временной упорядоченности между переменными и/или на основании теории.

[illegible] частной корреляции будет боль[illegible] $> r_{yx}$). Этот результат [illegible]

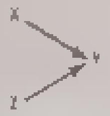
[illegible]

между переменными X и Z говорит об отсутствии между ними какой-либо связи. Эта [illegible]

[illegible]

лее сильное влияние на зависимую переменную. (См. задачи 17.1–17.3 по вычислению и интерпретации коэффициентов частной корреляции.)

Рис. 17.1. Возможная причинно-следственная [illegible]



17.3 Множественная регрессия: предсказание значений зависимой переменной

[illegible]

Формула 17.2 [illegible]

где: a — свободный член,
b — [illegible]

[illegible], определенную по методу наименьших квадратов, можно [illegible]менять, что позволяет включать (теоретически) любое [illegible] переменных. [illegible] множественной регрессии. Для простоты объяснения ограничимся случаем с двумя независимыми переменными [illegible] множественной регрессии [illegible] следующим образом:

Формула 17.3 [illegible]

где b_1 — частный коэффициент уравнения регрессии при первой независимой переменной;

b_2 — частный коэффициент уравнения регрессии при второй независимой [illegible]

В этой формуле введены новые условные обозначения и понятия. Зависимая переменная по-прежнему обозначается как Y, а вот независимые переменные различаются по подстрочному индексу. Так, X_1 определяет первую независимую переменную, а X_2 — вторую. Символ наклона линии регрессии b также имеет подстрочный индекс для обозначения связанной с ним независимой переменной.

Основное различие между множественной регрессией и парной регрессией заключается в частных коэффициентах уравнения регрессии b. В случае множественной [illegible] b называются частными коэффициентами регрессии [illegible] величину изменения Y, приходящегося на единичное изменение не[illegible] переменной при условии контроля над эффектами влияния других независимых переменных, входящих в уравнение. Таким образом, частные коэффициенты регрессии аналогичны коэффициентам частной корреляции и отражают прямой эффект влияния соответствующей независимой переменной на Y.

Частные коэффициенты уравнения регрессии при независимых переменных определяются в соответствии с формулами 17.4 и 17.5[1].

Формула 17.4 $$b_1 = \left(\frac{s_y}{s_1}\right)\left(\frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}\right)$$

Формула 17.5 $b_2 =$ [illegible]

где b_1 — [illegible]
b_2 — [illegible]
s_y — стандартное отклонение Y,
s_1 — стандартное отклонение первой независимой переменной (X_1),
s_2 — стандартное отклонение второй независимой переменной (X_2),
r_{y1} — коэффициент парной корреляции между Y и X_1,
r_{y2} — коэффициент парной корреляции между Y и X_2,
r_{12} — коэффициент парной корреляции между X_1 и X_2.

Чтобы показать схему вычисления частных коэффициентов регрессии, мы будем оценивать объединенное влияние количества детей в семье (X_1) и SES (X_2) на объем работ по дому, выполняемый главой семейства. Все необходимые данные можно вычислить из таблицы 17.1.

[1] [illegible]

[illegible]

Количество детей [illegible]

[illegible]

Коэффициенты корреляции нулевого порядка

$r_{yt} = 0{,}50$,

[illegible]

[illegible]

Частный коэффициент регрессии при первой независимой переменной (X_1)

$$b = \left(\frac{s_y}{s_1}\right)\left(\frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}\right)$$

[illegible]

$b = (1{,}4)(0{,}46)$,

$b_1 = 0{,}65$

Для второй независимой переменной (SES, или X_2), частный коэффициент регрессии будет таким:

[illegible]

$b_2 = (1{,}2\ldots)$ [illegible]

$$b_2 = (0{,}8)\left(\frac{-0{,}30+0{,}24}{0{,}78}\right)$$

[illegible]

$b = (0{,}8)(-0{,}08)$,

[illegible]

Теперь, когда нам известны частные коэффициенты регрессии при каждой независимой переменной, можно найти свободный член (a), или точку пересечения с осью Y. Обратите внимание на то, что a вычисляется из среднего значения зависимой переменной (обозначаемого $\bar{Y}$) и из средних значений двух независимых переменных [illegible]

Формула 17.6 $a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$

Подставив соответствующие значения в наш пример задачи, получим:

[illegible]

$a = 3{,}3 - (0{,}65)(2{,}7) - (-0{,}07)(13{,}7)$

$a = 3{,}3 - (1{,}8) - (-\ldots)$ [illegible]

[illegible]

Для нашего примера полное уравнение множественной регрессии, полученное по методу наименьших квадратов, будет иметь вид:

$Y = 2{,}5 + (0{,}65)X_1 + (-0{,}07)X_2$

Как и в случае с линейной парной регрессией, эту формулу можно использовать для предсказания значений зависимой переменной по значениям независимых переменных. Например, как можно наиболее точно предсказать объем работ по дому, выполняемый главой семейства (Y') для семьи с четырьмя детьми ($X_1 = 4$), если глава семейства потратил 11 лет на учебу в средней школе ($X_2 = 11$)? Подставив эти значения в формулу, получим:

[illegible]

[illegible]

регрессии, полученное по методу наименьших [illegible] информации относящуюся к независимым [illegible] гой метод предсказания [illegible] симой переменными существует [illegible] *предсказанию значений Y и вычисл*[illegible] *ных членов.*

17.4. Множественная регрессия: оценка влияния независимых переменных

Уравнение множественной регрессии, полученное по методу наименьших квадратов (формула 17.5) используется для обособления отдельных эффектов влияния независимых переменных и предсказания значений зависимой переменной. Однако во многих случаях использовать эту формулу для нахождения относительной значимости различных независимых переменных будет довольно тяжело, особенно в тех случаях, когда независимые переменные имеют различия в единицах измерения (например, количество детей в семье против количества лет учебы). У случаях, когда переменные имеют различные единицы измерения, сравнение частных коэффициентов регрессии не скажет нам ничего существенного о том, какая из независимых переменных оказывает наибольшее значительное влияние и является наиболее важной. Сравнение частных коэффициентов регрессии при переменных, имеющих различия в единицах измерения, подобно сравнению яблок и апельсинов.

И все-таки, сравнить независимые переменные можно, если привести все переменные в уравнении к одной общей шкале, что поможет избавиться от различия в [illegible]

[illegible] таблицу 5), а сравнение между независимыми переменными будут иметь смысл. Чтобы стандартизировать переменные к нормальной кривой, можно преобразовать все [illegible] затем повторно рассчитать частные коэффициенты [illegible]. Для этого понадобится выполнить большой [illegible] существует простой [illegible]

[illegible] ованные частные коэффициенты регрессии [illegible] коэффициент [illegible]

[illegible] значений каждой из независимых переменных при условии контроля над эффектами влияния всех остальных независимых переменных. Для двух независимых переменных бета-коэффициенты можно найти по формулам 17.7 и 17.8.

Формула 17.7 [illegible]

Формула 17.8

$$b_2^* = b_2\left(\frac{s_2}{s_y}\right)$$

где b_2^* — стандартизованный [illegible] коэффициент регрессии при [illegible]

Используя стандартизованные значения, уравнение регрессии, [illegible] методу наименьших квадратов можно переписать [illegible]

Формула 17.9 $Z_y = [illegible] + b_1^* Z_1 + b_2^* Z_2$

где: символ Z показывает, что все значения были стандартизованы к нормальной кривой.

Стандартизованное уравнение регрессии можно упростить, если убрать из него свободный член, поскольку он всегда будет равен нулю после стандартизации значений. Это значение соответствует точке, в которой поверхность, описываемая уравнением регрессии, пересекает ось Y, и равно среднему значению Y при равенстве нулю всех независимых переменных. Увидеть это соотношение можно, если в формуле 17.6 вместо всех независимых переменных подставить нулевые значения

$$[illegible]$$

$$= b_1(0) + b_2(0).$$

Так как среднее любого стандартизованного распределения значений равно нулю, то и среднее стандартизованных значений Y также будет равно нулю и, соответственно, точка пересечения с осью Y также будет соответствовать нулю ($a = \overline{Y} = 0$). Таким образом, формула 17.9 упрощается и принимает следующий вид.

Формула 17.10 $Z_y = b_1^* Z_1 + b_2^* Z_2$

Теперь в нашей задаче можно вычислить бета-коэффициенты, которые помогут определить, какая из двух независимых переменных оказывает наиболее сильное влияние на зависимую переменную. Для первой независимой переменной (количество детей в семье, или X_1):

$$b_1^* = b_1 [illegible]$$

$$b_1^* = (0.65)(0.7[illegible])$$

$$b_1^* = [illegible]$$

Для второй независимой переменной (SES, или X_2):

$$b_2^* = b_2 [illegible]$$

$$b_2^* = (-0.07)\left(\frac{2.6}{2.[illegible]}\right)$$

$$b_2^* = (0{,}07)([illegible],24).$$

[illegible]

Таким образом, после учета бета-коэффициентов стандартизованное уравнение регрессии примет в[illegible]

[illegible]

Из этого уравнения становится ясно, что первая независимая переменная [illegible] зывает более сильное влияние на Y, чем вторая независимая переменная.

В заключение следует отметить, что анализ множественной регрессии позволяет исследователю описать линейную связь между двумя или более независимыми переменными и зависимой переменной. Нестандартизованное уравнение регрессии (формула 17.2) позволяет предсказать значения Y по значениям независимых переменных в исходных единицах измерения значений переменных. Стандартизованное уравнение регрессии (формула 17.10) позволяет исследователю без особых [illegible] оценить относительную важность различных независимых переменных, сравни[illegible] [illegible] *претации бета-коэффициентов. Неплохо начать с решения задачи 17.1, поскольку* [illegible]

Чтение статистик 12: Регрессия и корреляция

В исследовательских проектах, [illegible] взаимосвязи между многими переменными, в качестве [illegible] методов практически всегда [illegible] ются методы регрессии и корреляции. [illegible] [illegible] корреляции, коэффициент [illegible] приводится объяснение [illegible] Пирсона для каждой [illegible] матрицы можно [illegible]

[illegible] коэффициента R [illegible] коэффициента регрессии для каждой независимой переменной, присутствующей в уравнении регрессии. Пример сводной таблицы:

Независимые переменные	Коэффициент R^2	Бета-коэффициент
X_1	0,17	0,47
X_2	0,[illegible]	0,[illegible]2
X_3	0,27	[illegible]

[illegible] таблицы [illegible] независимая переменная X [illegible] имеет [illegible] [illegible] независимой

переменной (R^2 = 0,17). Вторая независимая переменная (X_2) добавляет 6% в объясненную дисперсию (R^2 = 0,23 после ввода переменной X_2 в уравнение). Третья независимая [illegible] ной (R^2 = 0,27 после ввода переменной X_3 в уравнение).

Статистические данные в профессиональной литературе

Исследователи систематически публикуют данные о существовании связи между бедностью граждан и случаями убийств для белого населения, в то же время такая связь отсутствует для темнокожих представителей общества. Социолог Мэтью Ли (Matthew Lee) заинтересовался этим фактом и решил выяснить, связано ли это с тем, что при традиционных исследованиях не учитываются особенности пространственной концентрации малоимущего населения в центральных городах для представителей меньшинства. Иными словами, бедность городских афроамериканцев обусловлена сегрегацией населения и дискриминацией при предоставлении жилья. Действительно ли эта комбинация факторов является причиной высокого процента убийств? Ли собрал данные о случаях совершения убийств, неблагоприятных условиях жизни для темнокожего населения (сложная шкала, включающая меры безработицы, бедности, уровня жилищного комфорта и другие переменные) и данные по другим управляющим переменным для выборки 121 центрального города за 1990 год. Его результаты представлены в двух моделях, приведенных в табл. 1.

Таблица 1. Результаты множественной регрессии

	Модель 1		Модель 2	
	Нестандартизованный коэффициент	Бета-коэффициент	Нестандартизованный коэффициент	Бета-коэффициент
Неблагоприятные условия [illegible] граждан в городе	0,046	0,347	0,009	0,065
[illegible]			[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	
[illegible]	[illegible]	[illegible]		
Юг	-0,149	-0,172	-0,254	-0,226
[illegible]	[illegible]		[illegible]	
	[illegible]		[illegible]	

$^*p < 0{,}05$ $^{**}p < 0{,}01$

Адаптировано из табл. 3

В первой модели [illegible] эффект влияния неблагоприятных [illegible] темнокожих граждан в городе и управляющих переменных на процент [illegible] среди темнокожего населения: можно отметить положительную [illegible] мую связь. С ухудшением условий жизни возрастает [illegible] убийства для темнокожих граждан, даже после включения [illegible]

[illegible]

...ной концентрацией малоимущего населения...

[illegible]

свидетельствует о том, что связь в модели 1 является кажущейся и вызвана эффектами влияния пространственной концентрации бедности, а не просто бедностью или н... [illegible]

17.5. Множественная корреляция

Уравнения множественной регрессии мы используем для нахождения отдельных эффектов прямого влияния каждой независимой переменной на зависимую переменную. Использование метода **множественной корреляции** позволяет обнаружить объединенный эффект влияния всех независимых переменных на зависимую переменную. Это осуществляется посредством вычисления **коэффициента множественной корреляции (R) и коэффициента множественной детерминации (R^2)**. Значение [illegible]

[illegible] зависимыми переменными. Получив значения коэффициентов корреляции нулевого порядка, мы видели, что *количество детей в семье* (X_1) объясняет 0,25 дисперсии Y ($r^2_{y1} = (0{,}50)^2 = 0{,}25$), а SES объясняет 0,09 вариации Y ($r^2_{y2} = (-0{,}30)^2 = 0{,}09$) [illegible]

[illegible]

...нить, используя следующую формулу:

Формула 17.11 $R^2 = r^2_{y1} + r^2_{y2.1}(1 - r^2_{y1})$

где: R^2 — коэффициент множественной корреляции;

r^2_{y1} — корреляция нулевого порядка между Y и X_1, [illegible]

[illegible]

[illegible] Она представляет долю вариации Y, объясненную переменной X_1. К [illegible]

величине мы добавим долю оставшейся вариации в Y, выражаемую как $(1 - r^2_{y1})$, которая может быть объяснена переменной X_2 после установления контроля над эффектами влияния X_1 ($r^2_{y2.1}$). По существу, формула 17.11 позволяет вначале вычесть из влияния переменной X_1 на Y, а затем добавить эффект влияния переменной X_2 после установления контроля над X_1 (избавившись, тем самым, от "перекрытия" в вариации Y, которую совместно учитывают переменные X_1 и X_2).

Чтобы вычислить общий эффект влияния количества детей в семье X_1 и SES (X_2) на объем работ по дому, выполняемый главой семейства Y, нам понадобятся две величины. Корреляция между X_1 и Y нам уже известна ($r_{y1} = 0{,}50$), однако прежде чем произвести вычисления по формуле 17.11, мы должны еще вычислить частную корреляцию между Y и X_2 с учетом контроля над переменной X_1 ($r_{y2.1}$).

[illegible]

Теперь по этой формуле можно произвести вычисления для [illegible]

[illegible]

$$R^2 = (0{,}5)^2 + (-0{,}06)^2(1 - 0{,}5^2)$$

$$R^2 = 0{,}25 + (0{,}004)(1 - 0{,}25)$$

[illegible]

Первая независимая переменная (количество детей в семье, X_1) объясняет 25% вариации Y. К этому значению вторая независимая переменная (SES, X_2) добавляет только полпроцента, что в сумме дает объясненную вариацию 25,3%. Вместе обе [illegible] переменной. *В конце этой главы приводятся задачи по вычислению и интерпретации коэффициентов R и R^2. Неплохо начать с решения задачи 17.1, поскольку она содержит минимальный набор данных и не требует выполнения сложных вычислений.*

17.6 Интерпретация статистик: действительно ли размер компании имеет значение для ассоциированного управления?

[illegible] ...я переменная является логической управляющей переменной, по[illegible] ...ании, имеющие большое количество мест в правлении, потенциаль[illegible] ...ьшее количество ассоциированных директоров — простая функ[illegible] ...вовлеченных в руководство. Многомерный анализ области [illegible] прибыли корпорации и количества мест в правлении [illegible]

Таблица 17.2 Корреляция нулевого порядка

[illegible]	Прибыль за 1995 год	Число мест в правлении [illegible]
[illegible] правлении других корпораций	[illegible] (0.00)	[illegible] (0.00)
[illegible] за 1995 год		0.20

Как вам уже известно, данные для анализа были взяты из 125 случайно отобранных корпораций за 1995 год. Основные описательные статистики для общего коли[illegible]

Прежде чем перейти к выполнению многомерного анализа, следует провер[illegible] коэффициенты корреляции между всеми возможными парами переменных. В [illegible] Компании с большой годовой прибылью стремятся иметь больше связей бла[го]даря [illegible] правлении других корпораций, чем компании, имеющие меньшие годовые прибыли.

Переходя к рассмотрению следующей [illegible]

[illegible] "Corporate Interlocking in the United States, 1962–1995" [illegible] было представлено на ежегодном собрании Южного [illegible]

Как мы и предсказывали при выборе [illegible] имеющие большее количество мест [illegible] ей, чем компании, имеющие мен[ьше] [illegible] ...ров, занимающих места в ее правлении [illegible] ...мечается тенденция роста количества мест в правлении.

Перейдем к выполнению многомерного анализа и определим, существует ли первоначальная связь между общим количеством ассоциированных директоров и уровнем прибыли при условии контроля над эффектами влияния переменной "количество мест в правлении". Не забывайте о том, что коэффициент частной корреляции показывает силу и направление парной связи при условии контроля над эффектами влияния третьей переменной. В данном случае коэффициент корреляции нулевого порядка составлял 0,33, а коэффициент частной корреляции между общим количеством работающих по совместительству директоров и годовой прибылью корпорации, при условии контроля над количеством мест в правлении, составляет 0,29 ($r_{xy.z} = 0,29$). Так как величина частной корреляции лишь немного меньше величины исходной связи, можно прийти к заключению, что количество работающих по совместительству директоров и размер годовой прибыли корпорации будут иметь связь даже при условии контроля над количеством мест в правлении. Общее количество работающих по совместительству директоров возрастает с увеличением годовой прибыли вне зависимости от количества мест в правлении корпорации.

Таблица 17.3 Результаты множественной регрессии

	Модель 1			Модель 2		
	Нестандартизованный коэффициент	Бета коэффициент	t значение	Нестандартизованный коэффициент	Бета коэффициент	t значение
Свободный член (a)	0.19	0.1[illegible]	0.01	0.0[illegible]		
Количество мест [illegible]	[illegible]	[illegible]	[illegible]	0.[illegible]	0.30	[illegible]
Прибыль [illegible]				[illegible]	[illegible]	[illegible]
R^2 (коэффициент множественной детерминации)	0.13			0.20		

[illegible]

[illegible] сетно мест в правлении используется для предсказания общего [illegible] совместительству директоров. Во второй модели [illegible] независимая переменная) и, что более важно для оценки [illegible] второй независимой переменной, свидетельствующий, что [illegible] несколькими тесно связанными между собой корпорациями [illegible] модели, можно оценить относительную важность прибыли [illegible] в предсказании общего количества работающих [illegible] (модель 2), а также дополнительную степень объяснения [illegible] количества мест в правлении.

[illegible] количество работающих по совместительству директоров [illegible] (*b* или *нестандартизованные коэффициенты*) является [illegible] мест в правлении объясняет 13% дисперсии общего количества [illegible]

[illegible] регрессии или частные коэффициенты регрессии *b*. Они, как вам уже известно, показывают изменение зависимой переменной, порождаемое [illegible] переменной при условии контроля над остальными независимыми переменными. Эти коэффициенты можно подставить в уравнение множественной регрессии (см. формулу 17.3) и использовать для предсказания значений зависимой переменной при различных комбинациях значений независимых переменных. Например, какое количество директоров-членов правления других корпораций мы сможем предсказать для корпорации, в правлении которой насчитывается 10 человек X_1 и годовая прибыль которой составляет 15 миллионов долларов X_2?

$$Y = a + b_1X_1 + b_2X_2$$

[illegible]

В модели 2 также представлены стандартизованные коэффициенты регрессии [illegible] бета-коэффициенты [illegible]

Бета-коэффициенты показывают, что эффекты влияния двух независимых [illegible]

[illegible]

[illegible] R^2 между двумя моделями. В модели 1 количество мест в правлении учитывает 13% вариации [illegible]

[illegible] *дополнительные* 7% объяснения, добавив переменную прибыли.

Напоследок давайте подытожим результаты, полученные в ходе решения задачи. Коэффициент частной корреляции между [illegible] совместительству директоров и прибылью при условии контроля над переменной *количество мест в правлении* остается довольно значительным. Анализ множественной регрессии показывает, что прибыль и количество мест в правлении имеют важные [illegible] изменение значений R^2 от модели 1 к модели 2, мы увидим дополнительные 7% объяснения. Поэтому даже с учетом влияния переменной *количество мест в правлении* размер (измеряемый прибылью) *действительно будет иметь значение* для расширения связей корпораций, которое осуществляется путем предоставления мест в правлении директорам-совместителям.

Практический пример 17.1

Пятерых мужчин, расторгнувших в недавнем времени брак, попросили дать субъективную оценку того, насколько удачно они смогли приспособиться к холостяцкой жизни. Для оценки была предложена пятибалльная шкала, в которой балл 5 соответствовал очень хорошей приспособленности, а балл 1 — очень плохой. Действительно ли степень приспособленности связана с продолжительностью брака? Верно ли, что степень приспособленности связана с социально-экономическим статусом, измеряемым по уровню первоначального годового дохода?

Наблюдение	Степень приспособленности (Y)	Продолжительность брака (X_1)	Годовой доход (в долларах) (X_2)
A	5	5	300000
B	4	7	45000
C	4	0	25000
D	3	2	27000
E	[illegible]	15	17000
	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]173,85

[illegible]

	Продолжительность брака (X_1)	Годовой доход (в долларах) (X_2)
Степень приспособленности (Y)	-0,62	0,62
[illegible]		[illegible]

Эти результаты свидетельствуют о сильной [illegible] отрицательной связи между каждой независимой переменной и переменной *степень приспособленности*. Степень приспособленности снижается с увеличением количества лет, прожитых в браке, и возрастает с повышением уровня годового дохода.

Чтобы получить уравнение множественной регрессии, необходимо вычислить частные коэффициенты регрессии.

Для количества лет, прожитых в браке (X_1)

[illegible]

$$b_1 = (0,3[illegible])(-0,42)$$

$$b_1 = -0,13$$

Частный коэффициент регрессии для [illegible]

[illegible]

$$b_2 = \frac{4}{9173{,}88}\left(\frac{(0{,}62) - (0{,}62)(-0{,}49)}{1-(-0{,}49)^2}\right),$$

[illegible]

$$b_2 = (0{,}000[illegible]5)\left(\frac{0{,}32}{0{,}76}\right)$$

$$b_2 = (0{,}000[illegible]5)(0{,}42)$$

b_2 [illegible]

Свободный член

[illegible]

$a = 2,60$

[illegible] множественной регрессии

[illegible]

$$Y' = 2{,}60 + (-0{,}[illegible]3)X_1 + (0{,}000063)X_2$$

Какое значение степени приспособленности [illegible] для мужчины, прожившего в браке 30 лет (X_1 [illegible]), годовой доход которого составляет 50000 долларов ($X_2 = 50000$)?

$$Y' = 2{,}60 + (-0{,}13)(30) + (0{,}000063)(50000);$$

$$Y' = 2{,}60 + (-3{,}9) + (3{,}15),$$

$$Y' = 1{,}85$$

[illegible] переменная оказывает наиболее сильное влияние на степень приспособленности, необходимо вычислить стандартизованные частные коэффициенты регрессии. Для продолжительности брака (X_1):

[illegible]

[illegible]

b_1^* [illegible]

[illegible]

$$b_2^* = b_2\left([illegible]\right)$$

[illegible]

$b_2^* = 0{,}41$

Стандартизованное уравнение регрессии:

$$Z_y = b^*_1 Z_1 + b^*_2 Z_2 = (-0{,}42)Z_1 + (0{,}41)Z_2$$

[illegible] сти. Чтобы оценить объединенный эффект влияния двух независимых переменных на степень приспособленности, необходимо вычислить коэффициент множественной детерминации

$$R^2 = [illegible]$$

$$R^2 = (-0{,}62)^2 + (0{,}46 [illegible]$$

$$R^2 = 0{,}38 + (0{,}21)(1 - 0{,}38)$$

$$R^2 = [illegible]$$

$$R^2 = 0{,}38 + 0{,}13;$$

$$R^2 = 0{,}51$$

Первая независимая переменная *продолжительность брака* объясняет 38% вариации степени приспособленности. Годовой доход объясняет дополнительные 13% вариации, а вместе обе переменные объясняют в общей сложности 51% дисперсии степени приспособленности.

17.7. Ограничения применимости методов множественной регрессии и корреляции

[illegible]

[illegible] дешевые. Для них требуются данные высокого качества, а наблюдение переменных, принадлежащих [illegible] шкале, является достаточно сложным [illegible] одолги. Более того, [illegible]

[illegible] либо взаимодействие ме[illegible]

дать точную оценку зависимой переменной [illegible] сложив эффекты влияния независимых переменных. [illegible]

[illegible]

[illegible] действительность, коэффициенты [illegible] и коэффициент множественной [illegible] необходимо too обрать альтернатив[illegible] речь в предыдущей главе. Однако они, к сожалению, могут [illegible]ожить лишь ограниченное количество менее точной информации о зависимостях между переменными.

В завершение следует отметить, что мы рассмотрели только самые простые случаи применения частной корреляции, множественной регрессии и корреляции. В смысле логики и интерпретации эти методы можно относительно просто применить в ситуациях, в которых участвуют несколько управляющих переменных или более двух независимых переменных. Однако вычисления в подобных случаях являются крайне сложными. Если вы столкнетесь с ситуацией, когда в анализ будут включены более трех переменных, воспользуйтесь одним из статистических пакетов, установленных в вашем высшем учебном заведении (например, SPSS и MicroCase). Для работы с этими программами вам понадобятся элементарные навыки работы на персональном компьютере, эти программы способны выполнять сложные вычисления буквально в одно мгновение. Умелое использование этих пакетов позволит избавиться от монотонной работы и предоставит возможность делать то, что по душе [illegible] каждому социологу: размышлять о важности результатов и сущности общественной жизни.

[illegible]

Профессор Сиу Квонг Вонг (Siu Kwong Wong) недавно провел исследование преступности [illegible] китайской канадской молодежи [illegible] Виннипеге, Манитоба. Изучив существующие теории преступности, профессор Вонг задался вопросом [illegible] является ли фактор ассимиляции [illegible] социальной группе? Для китайской культуры, как часто считают, характерна ортодоксальность взглядов [illegible] уровень преступности [illegible] ственен более высокий уровень преступности? Судя по данным, предст[illegible]

[illegible]

...ые проступки и серьезные преступления [illegible] ...его варианта. Помимо этого, наличие друзей-[illegible]

[illegible] в результате профессор [illegible] пришел к заключению, [illegible] в данной группе является не фактор культурной [illegible]

[illegible]

Таблица А. Регрессия уровня преступности от роста культурного уровня

Независимая переменная	Бета [illegible]	Серьезные [illegible]	Незначительные проступки
Пол	-0,25	-0,11	[illegible]
	[illegible]	[illegible]	
[illegible] роста культурного уровня	0,21	0,10	[illegible]
N	285	285	[illegible]
R^2	0,17	0,04	[illegible]

[illegible] в уравнении регрессии переменной, относящейся к [illegible] ...на две категории. В данном случае пол имеет две категории: 1 — мужчины и 2 — [illegible]

Заключение

1 Частная корреляция включает контроль над третьими переменными, [illegible] ...позволяет выявлять ранее кажущиеся или опосредованные [illegible]

2 [illegible]

3 [illegible] ...взаимосвязь с временными.

4 Частная корреляция, множественная регрессия и корреляция являются одними [illegible] требуется проведение высококачественных измерений и наличие связей между переменными, являющихся линейными и взаимосвязей между ними. Кроме этого, корреляции между независимыми переменными должны иметь малые значения (желательно нулевые). Несмотря на высокую стоимость, эти методы позволяют достичь желаемой точности и получить детальную информацию о взаимосвязях между переменными.

Сводка формул

[illegible]

17.1 [illegible]

Линия регрессии по методу наименьших квадратов (двумерная)

17.2 $Y = a + bX$

Линия множественной регрессии по методу наименьших квадратов

17.3 $Y = a + b_1X_1 + b_2X_2$

Частный коэффициент регрессии для X_1

17.4 [illegible]

Частный коэффициент [illegible]

[illegible]

[illegible] (бета-коэффициент) для X_1 [illegible]

17.7 [illegible]

[illegible] (бета-коэффициент) для X [illegible]

17.8 [illegible]

[illegible] построенная по методу наименьших квадратов.

[illegible]

Стандартизованная линия регрессии, построенная по методу наименьших квадратов (упрощенный вариант).

[illegible]

Коэффициент множественной детерминации

17.[illegible] $R^2 = r_{y1}^2 + r_{y2.1}^2 (1 - r_{y1}^2)$

Словарь

Бета-коэффициенты (b^*). Стандартизованные частные коэффициенты регрессии [illegible]

Коэффициент множественной детерминации (R^2). Статистика, равная общей по [illegible]

Коэффициент множественной корреляции (R). [illegible] ту связи между зависимой переменной и двумя или более независимыми пе[illegible]

Коэффициент частной корреляции. [illegible] ду двумя переменными при условии контроля над остальными переменными [illegible]

Множественная корреляция. Многомерный метод проверки объединенного эф[illegible] та влияния нескольких независимых переменных на зависимую перемен[illegible]

Множественная регрессия. Многомерный метод, позволяющий анализ [illegible] переменных на зависимую пер[illegible] ную; используется для предсказания [illegible]

Стандартизованные частные коэффициенты регрессии (бета-коэффициенты). [illegible] переменной после стандартизации всех значений.

Частная корреляция. Многомерный метод проверки парной связи при условии контроля над другими переменными.

Частный коэффициент регрессии. Выражает изме[illegible] переменной при [illegible] переменной [illegible] других [illegible] переменных [illegible] оценки множественной регрессии остаются неизменными.

Мультимедийные ресурсы

The Wadsworth Sociology Resource Center: Virtual Society (Виртуальное сообщество Центра социологических ресурсов Вадсворта)

http://sociology.wadsworth.com/

Посетите web-сайт шестого издания книги *Statistics: A Tool for Social Research*, где вы сможете получить доступ к широкому диапазону студенческих ресурсов. Щелкните в разделе Student Resources [illegible]

- Обзор основ математики
- Обзор статистики
- Основные [illegible] карты
- Интернет ссылки
- Дополнительные задачи к главам
- Таблица случайных чисел
- Примеры и упражнения с применением пакетов статистической обработки данных MicroCase и *SPSS*
- Функциональные диаграммы "Find the test"
- Проверка гипотез для переменных, измеренных по порядковой шкале

Задачи

17.1. PS

В [illegible] представлены данные об отказе избирателями в поддержке [illegible]

Город	Отказ в поддержке	Уровень безработицы	Процент [illegible]
[illegible]	55	5	[illegible]
B	60	8	63
[illegible]	65	7	[illegible]
D	66	9	[illegible]
E	[illegible]	[illegible]	[illegible]

Одномерные статистики.

	Отказ в поддержке	Уровень безработицы	Процент антирекламы
	63,8	8,2	55,8
[illegible]	5,5	1,7	5,3

Парные корреляции

	Уровень безработицы	Процент антирекламы
[illegible]	0,95	-0,87
[illegible]		[illegible]

а. [illegible]

б. Вычислите коэффициент частной корреляции для связи между отказом в поддержке Y и антирекламой X при условии контроля над переменной *уровень безработицы* Z. Каким образом влияет управляющая переменная на парную связь? Действительно ли является прямой связь между отказом в [illegible] 17.1 [illegible] *дел 17.2. Частная корреляция понадобится для нахождения коэффициента* [illegible]

в. Составьте нестандартизованное уравнение множественной регрессии, в котором *уровень безработицы* X_1 и *процент антирекламы* X_2 будут выступать в качестве независимых переменных. Какой процент отказа в поддержке следует ожидать в городе, в котором уровень безработицы составляет 10% [illegible] *мулу 17.6 для нахождения свободного члена а (пересечения с осью Y). [illegible] регрессии можно определить по формуле 17.1. Чтобы вычислить предсказ*[illegible]

г. Вычислите бета-коэффициенты для каждой независимой переменной. Какая из них оказывает наиболее существенное влияние на переменную [illegible] *те формулы 17.7 и 17.8.*

д. Вычислите коэффициент множественной корреляции R и коэффициент множественной детерминации R^2. Какая доля изменчивости перемен[illegible] *сказка: используйте формулу 17.11. Значение* r^2_{YZ} *вы вычислили в части б* [illegible] задачи.

е. Опишите связи между этими тремя переменными.

17.2. SOC

В выборке представлены данные о голосах поддержки в пользу увеличения государственных расх[illegible] о количестве лет учебы и о количестве лет службы в армии (если они проходили службу в армии). [illegible]

Наблюдение	Поддерживаю	Количество лет учебы	Количество лет службы в армии
[illegible]	[illegible]	[illegible]	2
В	15	12	4
[illegible]	20	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	0
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	5	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
Н	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

а. Вычислите коэффициент частной корреляции для связи между поддержкой Y и количеством лет учебы X при условии контроля над количеством лет службы в армии Z. Каким образом влияет управляющая переменная на парную связь? Действительно ли является прямой связь между поддержкой и количеством лет учебы?

б. Вычислите коэффициент частной корреляции для связи между поддержкой Y и количеством лет службы в армии X при условии контроля над количеством лет учебы Z. Действительно ли является прямой связь между поддержкой и количеством лет службы в армии? *Подсказка: частная корреляция необходима для вычисления коэффициента множественной корреляции.*

в. Получите нестандартизованное уравнение множественной регрессии, в котором переменная "количество лет учебы" X_1 и переменная *количество лет службы в армии* X_2 выступали бы в качестве независимых переменных. Какую степень поддержки можно было бы ожидать от лиц, проучившихся 13 лет и прослуживших 15 лет в армии?

г. Вычислите бета-коэффициенты для каждой независимой переменной. Какая из них оказывает наиболее существенное влияние на отказ в поддержке?

д. Вычислите коэффициент множественной корреляции R и коэффициент множественной детерминации R^2. Какая доля изменчивости переменной *поддержка* объясняется двумя независимыми переменными? *Подсказка:* [illegible]

е. Опишите связи между этими тремя переменными.

17.3. SOC

[illegible]

[illegible] переменную *межусобица*. Вычислите корреляции [illegible] между тремя переменными.

[illegible]	Уровень безработицы	Процент населения, проживающего в городских районах
0	5,3	66
1	[illegible]0	65
5	2,7	22
[illegible]	2,6	86
[illegible]	3,0	69
[illegible]	2,5	76
[illegible]	6,0	45
[illegible]	5,2	46
[illegible]	7,8	75
[illegible]	9,2	80

а. Вычислите коэффициент частной корреляции для связи между переменными *межусобица Y* и *уровень безработицы X* при условии контроля над [illegible] [illegible] безработицей.

б. [illegible] бицей Y и урбанизацией X при условии контроля над безработицей Z. Каким образом влияет управляющая переменная на парную связь? Действительно ли является прямой связь между межусобицей и урбанизацией [illegible] *енты множественной корреляции*.

в. [illegible] сказать для государства, в котором уровень безработицы составляет [illegible]0%, 90% населения проживает в городских районах.

г. Вычислите бета-коэффициенты для каждой независимой переменной. Какая из них оказывает наиболее существенное влияние на межусобицу?

д. [illegible] *межусобица* объясняется двумя независимыми переменными?

е. [illegible]

17.4. [illegible]

[illegible]

а. [illegible]

б. [illegible]

в. Вычислите бета-коэффициенты для каждой независимой переменной [illegible] сравните их относительные эффекты влияния на каждую зависимую переменную.

г. [illegible] R [illegible] в качестве независимых переменных.

д. Опишите полученные результаты.

17.5. PS

[illegible] *поддержке*.

а. [illegible]

б. [illegible] ми демократической партии, а 5% — сторонники партии, находящейся в меньшинстве.

в. Вычислите бета-коэффициенты для каждой независимой переменной и сравните их относительные эффекты влияния на отказ в поддержке. Какая из них оказывает наиболее существенное влияние?

г. Вычислите коэффициенты R и R^2.

д. Опишите полученные результаты.

17.6. SW

Двенадцать семей обратились к консультанту по брачно-семейным отношениям. Консультант составил таблицу, в которую внесли данные о семьях, используя [illegible] информацией о ежегодном доходе семьи и количестве детей, проживающих в настоящее время вместе с семьей. В качестве зависимой переменной используется [illegible] *зованность членов семьи*.

Семья	[illegible] членов семьи	Ежегодный доход семьи	Количество детей
[illegible]	[illegible]	[illegible]	[illegible]

Семья	[illegible] членов семьи	Ежегодный доход семьи	Количество детей
		[illegible]000	[illegible]
[illegible]		[illegible]000	6
	4		

[illegible] множественной регрессии

а. [illegible] можно предсказать для семьи [illegible] 5 [illegible] 6 детей

в. [illegible] коэффициенты для каждой независимой переменной и [illegible]. Какая из них оказывает наиболее существенное влияние?

г. Вычислите коэффициенты R и R^2.

д. Опишите полученные результаты.

[illegible] В задаче 15.[illegible] были представлены данные о расходах на образование на душу населения для [illegible]5 штатов, ранге ежегодного дохода на душу населения и проценте лиц, имеющих аттестат об окончании средней школы. В качестве зависимой переменной используйте *расходы на образование*.

а. Вычислите бета-коэффициенты для каждой независимой переменной и сравните их относительные эффекты влияния на переменную *расходы на образование*. Какая из них оказывает наиболее существенное влияние?

б. Вычислите коэффициенты R и R^2.

в. Опишите полученные результаты.

17.8 [illegible]

Из базы данных всеобщего социального обследования (General Social Survey, GSS) за 1998 год были отобраны данные для 20 индивидуумов, значения [illegible] четырех переменных представлены далее в таблице. Используйте *количество часов, потраченных на просмотр телевизионных программ в течение дня* в качестве [illegible] переменной, а две из оставшихся — в качестве независимых [illegible]

Количество часов, потраченных на просмотр телевизионных программ в течение дня (TVHOURS)	Фактор престижа работы (PRESTG80)	Количество детей (CHILDS)	Возраст AGE
4	50	2	[illegible]
[illegible]	[illegible]	3	[illegible]
3	36	4	[illegible]
4	50	[illegible]	[illegible]
2	45	[illegible]	[illegible]
3	50	[illegible]	[illegible]
4	50	[illegible]	[illegible]
7	[illegible]	3	[illegible]
[illegible]	57	[illegible]	[illegible]
5	[illegible]	2	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Количество часов, потраченных на просмотр телевизионных программ в течение дня (TVHOURS)	Фактор престижа работы (PRESTG80)	Количество детей (CHILDS)	Возраст AGE
3	[illegible]	[illegible]	[illegible]
[illegible]	9	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
4	36	1	24
9	48	[illegible]	25
1	[illegible]	1	[illegible]
5	50	0	45
1	27	3	62

а. Вычислите бета-коэффициенты для каждой выбранной вами независимой переменной и сравните их относительные эффекты влияния на зависимую переменную. Какая из них оказывает наиболее существенное влияние?

б. Вычислите коэффициенты R и R^2.

в. Опишите полученные результаты.

SPSS for Windows

Использование *SPSS for Windows* для регрессионного анализа

Демонстрационный пример с использованием SPSS 17.1. Какие факторы коррелируют с фактором престижа работы? Другой взгляд на вопрос

В демонстрационном примере с использованием SPSS 15.[illegible] мы использовали процедуру **Correlation** (Корреляция) для вычисления коэффициентов корреляции нулевого порядка между переменными *prestg80*, *papres80* и *educ*. В этом разделе мы будем использовать более сложную и гибкую процедуру **Regression** (Регрессия), позволяющую анализировать эффекты влияния переменных *papres80* и *educ* на переменную *prestg80*. Процедура **Regression** позволяет управлять многими аспектами формулы регрессии, причем объем информации, которую можно получить в ходе выполнения данной процедуры, будет существенно больше того объема данных, который можно получить при выполнении процедуры **Correlation**. Помимо всего прочего, процедура **Regression** может показать коэффициент наклона b и свободный член a [illegible] точку пересечения с осью Y, поэтому ее можно использовать для нахождения линии регрессии [illegible]. В [illegible] примере [illegible] использование всех доступных опций. Есть смысл самостоятельно изучить некоторые варианты и возможности этой мощной процедуры анализа данных.

[illegible] GSS за 1998 [illegible] Analyze [illegible] Regression [illegible] Linear [illegible] Dependent [illegible] Linear Regression [illegible]

[illegible] Independent(s) (Независимые переменные) [illegible] Descriptives (Описательные) [illegible] Statistics [illegible] Continue [illegible] OK [illegible]

Итоговая сводка по модели

Модель	R	R-квадрат	Скорректированный R-квадрат	Стандартизованная ошибка оценки
[illegible]	0,518	0,269	0,267	11,78

Таблица дисперсионного анализа

Модель	Сумма квадратов	Ст. св.	Средний квадрат	F	Значимость
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Коэффициенты

Модель	Нестандарт. коэффициенты	Стандарт. ошибки	Стандарт. коэффициенты бета	t	Значимость
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
уровень [illegible]	[illegible]	[illegible]	0,[illegible]40	[illegible]	[illegible]
законченных классов	[illegible]	[illegible]	0,[illegible]25	[illegible]	[illegible]

[illegible] множественной корреляции R, которое составило 0,518, и значение коэффициента множественной детерминации R-квадрат, которое составило 0,269. В таблице результатов дисперсионного анализа показан уровень значимости связи (Знач. = 0,000). Мы знаем, что независимые переменные объясняют около 27% изменчивости переменной *prestg80* и что этот результат является статистически значимым. В последнем блоке указаны коэффициенты регрессии B при независимых переменных [illegible]

О чем можно судить на основании полученных результатов? [illegible]

Демонстрационный пример с использованием SPSS 17.2. Какие факторы коррелируют с сексуальной активностью?

[illegible] от экономического положения?

[illegible] переменные, имеющие всего две категории, — называемые [illegible] ременными (dummy) — довольно часто используются [illegible] ные статистики 13"). Нажмите команду Analyze (Анализ), Regression [illegible] Linear (Линейная) и объявите переменную *sexfreq* в качестве зависимой [illegible] описательные статистики, нажмите кнопку Statistics (Статистики) [illegible] ите соответствующий выбор. В результате на экран будет выведена информация, подобная следующей (описательная информация о переменных и корреляции нулевого порядка здесь не приводится с целью экономии места).

Итоговая сводка по модели

Модель	R	R-квадрат	Скорректированный R-квадрат	Стандартизованная ошибка оценки
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Таблица дисперсионного анализа

Модель	Сумма квадратов	Ст. св.	Средний квадрат	F	Значимость
Регрессия	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Остаток	[illegible]	[illegible]	[illegible]		
Итого	[illegible]	[illegible]			

Коэффициенты

Модель	Нестандарт. коэффициенты	Стандарт. ошибка	Стандарт. коэффициенты бета	t	Значимость
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Возраст [illegible]	[illegible]E-02	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]E-02	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

... сексуальных отношений и только за ним следуют ежегодный доход и ... [illegible]. Интенсивность сексуальных отношений возрастает с увеличением ... [illegible], уменьшается с ростом возраста и более связана с мужчинами, чем с [illegible]

Упражнения

17.1 Выполните анализ, предложенный в демонстрационном примере с использованием SPSS либо MicroCase 17.1, используя переменную *income91* в качестве зависимой переменной. Сравните результаты с результатами демонстрационного примера с использованием SPSS либо MicroCase 15.1. Объясните полученные [illegible]

17.2 Выполните анализ, предложенный в демонстрационном примере с использованием SPSS либо MicroCase 17.2, используя переменную *partnrs5* (количество сексуальных партнеров за последние пять лет) в качестве зависимой переменной. Сравните и сопоставьте полученные результаты с результатами анализа переменной *sexfreq*.

17.3 Выполните регрессионный анализ с участием переменных *tvhours*, *childs* или *educ*. Выберите две или три потенциальные независимые переменные и [illegible] MicroCase 17.1 и 17.2. В идеале ваши независимые переменные должны иметь [illegible] ... переменные с широким диапазоном значений и номинальные переменные, имеющие только два значения.

Итоговые упражнения к IV части

1. Далее сформулированы две задачи [illegible] ... мере должны соответствовать шкалы измерения переменных.

   a. [illegible] ... ее часов работы и среднего балла (в шкале единиц, ... [illegible] ... в качестве зависимой переменной используйте средний балл отметок.

   b. Только около половины студентов в этой выборке [illegible] (1 = да, 2 = нет). Оказывала ли ... на успехи в учебе активность (1 = низкая, 2 = высокая) студентов? Одинакова ли связь между этими переменными для юношей (1) и девушек (2)?

Средний балл [illegible]	Количество рабочих часов [illegible]	Средний балл [illegible]	Физическая [illegible]	[illegible]
3.14	13	650	1	1

2. В исследовании [illegible] [illegible]

[illegible]

ных программ?

2. Нет

К какой религиозной конфессии Вы относитесь?

1. [illegible]

2. Католик

Ваш пол

1. Мужской

2. Женский

в. Количество времени на просмотр телевизионных программ на религиозную тематику будет увеличиваться с ростом возраста и уменьшаться с повышением уровня образования, однако возраст будет оказывать на [illegible]более сильное влияние.

г. Женщины, вероятно, будут верить [illegible] всего протестанты. Женщины-протестантки будут особенно [illegible] предрасположены к [illegible]

Действительно ли данные, приведенные в следующей таблице, подтверждают [illegible]

[illegible]	[illegible]	Возраст	Образование	[illegible]	Религиозная конфессия	Пол
1	14	52	10	1	1	1
2	10	45	12	1	1	
3	6	18	12	1	1	
4	5	45	12	1	1	
					1	
6	14	65	8	2		
7	3	23	12	1	1	
8	12	47	11	2	1	
9	[illegible]	[illegible]	[illegible]	[illegible]	1	
10	1	20	16	2	1	
13	12	47	9	2	2	
14	8	32	14	1	2	
[illegible]	6	50	12	1	2	
16	10	45	10	[illegible]	2	
	12	40	16	1	2	
18	14	42	14	2	2	
[illegible]				2	2	

А Площадь под кривой нормального распределения

[illegible] Z [illegible] общей площади между текущ[illegible] Z [illegible] Z, а на рис. А.2 показан [illegible]

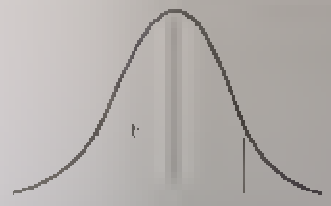


Рис. А.1. [illegible]

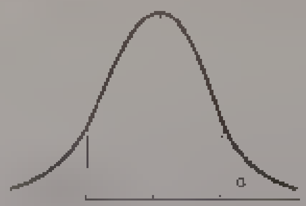


Рис. А.2. [illegible]

(а) Z	(б) Площадь между средним и Z	(в) Площадь за пределами Z
[illegible]	[illegible]	[illegible]

(а) Z	(б) Площадь между средним и Z	(в) Площадь за пределами Z
[illegible]	[illegible]	[illegible]

(a) Z	(b) Площадь между средним и Z	(c) Площадь за пределами Z
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0.4090
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0.3859
[illegible]	[illegible]	[illegible]
0,31	0.1217	0.3783
[illegible]	0.1255	0.3745
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0.3632
[illegible]	[illegible]	0.3594
0,37	0.1443	0,3557
0.38	0,1480	0,3520
[illegible]	[illegible]	0.3483
0.40	0.1554	0,3446
0,41	0.1591	0,3409
0,42	0,1628	0,3372
[illegible]	[illegible]	[illegible]
0.44	0,1700	0,3300
0.45	0.1736	0.3264
0.46	0.1772	0.3228
0,47	0.1808	0,3192
0.48	0.1844	0,3156
0.49	0,1879	0,3121
0.50	0.1915	0.3085
[illegible]	[illegible]	0.3050
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
0,54	0,2054	0,2946
0,55	0.2088	0,2912
0,56	0,2123	0,2877
0,57	0,2157	0,2843
[illegible]	[illegible]	[illegible]
[illegible]	0.2224	0.2776
[illegible]	0.2257	[illegible]

(a) Z	(b) Площадь между средним и Z	(c) Площадь за пределами Z
0.61	[illegible]	0.2709
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
0.65	[illegible]	[illegible]
[illegible]	0.2454	0.2546
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
0.69	0.2549	[illegible]
0,70	0,2580	0,2420
0.71	0,2611	0,2389
[illegible]	0.2642	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
0.76	0,2764	0,2236
0.77	0.2794	0,2206
0.78	0.2823	0.2177
0.79	0,2852	0.2148
0.80	0.2881	0,2119
0,81	0,2910	0,2090
[illegible]	[illegible]	[illegible]
0,83	0,2967	0.2033
0.84	0.2995	0.2005
0.85	[illegible]	[illegible]
0.86	0.3051	0.1949
0.87	0.3078	0.1922
0.88	0.3106	0.1894
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
0.93	0,3238	0,1762
0.94	0,3264	0.1736
0.95	0.3289	0.1711
[illegible]	0.3315	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0.1635
[illegible]	[illegible]	[illegible]

(a) Z	(b) Площадь между средним и Z	(c) Площадь за пределами Z
1.00	[illegible]	[illegible]
1.01	0.3438	0.1562
1.02	0.3461	0.1539
[illegible]	0.3485	[illegible]
1.04	0.3508	[illegible]
[illegible]	0.3531	[illegible]
[illegible]	0.3554	[illegible]
[illegible]	0.3577	[illegible]
[illegible]	0.3599	[illegible]
[illegible]	[illegible]	[illegible]
1.10	0.3643	0.1357
[illegible]	0.3665	0.1335
[illegible]	0.3686	[illegible]
[illegible]	0.3708	[illegible]
[illegible]	[illegible]	[illegible]
1,15	0,3749	0.1251
1,16	0.3770	0,1230
[illegible]	0.3790	[illegible]
1.18	0.3810	0,1190
1.19	0,3830	0.1170
1.20	0.3849	0.1151
[illegible]	0.3869	[illegible]
1.22	0,3888	0,1112
1,23	0.3907	0.1093
[illegible]	0,3925	[illegible]
1.25	0.3944	[illegible]
1.26	0.3962	0.1038
1.27	0.3980	0.1020
[illegible]	[illegible]	[illegible]
[illegible]	0.4015	0.0985
[illegible]	0.4032	0.0968
1.31	0.4049	0.0951
1.32	0.4066	0.0934
1.33	0.4082	[illegible]
1.34	0.4099	[illegible]
[illegible]	[illegible]	0.0885
[illegible]	[illegible]	0.0869
[illegible]	[illegible]	0.0853

(a) Z	(b) Площадь между средним и Z	(c) Площадь за пределами Z
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
1.40	[illegible]	0.0808
[illegible]	0.4207	[illegible]
[illegible]	0.4222	[illegible]
[illegible]	0.4236	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4265	[illegible]
[illegible]	[illegible]	[illegible]
1,47	[illegible]	[illegible]
1,48	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4345	[illegible]
[illegible]	[illegible]	0.0643
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4406	0.0594
[illegible]	0.4418	0.0582
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4452	0.0548
[illegible]	0.4463	0.0537
[illegible]	0.4474	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4505	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4525	[illegible]
[illegible]	0.4535	[illegible]
[illegible]	0.4545	0.0455
[illegible]	[illegible]	[illegible]
[illegible]	0.4564	0.0436
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4599	[illegible]

(a) Z	(b) Площадь между средним и Z	(c) Площадь за пределами Z
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0,4693	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
1,94	0,4738	0,0262
[illegible]	[illegible]	[illegible]
[illegible]	0,4750	[illegible]
[illegible]	[illegible]	[illegible]
1,98	0,4761	0,0239
[illegible]	[illegible]	[illegible]
2,00	0,4772	0,0228
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
2,06	0,4803	0,0197
2,07	0,4808	0,0192
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
2,10	0,4821	0,0179
2,11	0,4826	0,0174
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

(a) Z	(b) Площадь между средним и Z	(c) Площадь за пределами Z
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0,0158
[illegible]	[illegible]	0,0154
[illegible]	[illegible]	0,0150
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0,0110
[illegible]	[illegible]	0,0107
[illegible]	[illegible]	[illegible]
2,32	0,4898	0,0102
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
2,36	0,4909	0,0091
[illegible]	[illegible]	[illegible]
2,38	0,4913	0,0087
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
2,44	0,4927	0,0073
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0,0068
2,48	0,4934	0,0066
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

(a) Z	(b) Площадь между средним и Z	(c) Площадь за пределами Z
[illegible]	[illegible]	0,0059
2,53	0,4943	0,0057
2,54	0,4945	0,0055
2,55	0,4946	0,0054
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
2,66	0,4961	0,0039
2,67	0,4962	0,0038
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
2,70	0,4965	0,0035
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
2,74	0,4969	0,0031
[illegible]	[illegible]	[illegible]
2,76	0,4971	0,0029
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0,4975	[illegible]
2,82	0,4976	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
2,85	0,4978	[illegible]
2,86	0,4979	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

(a) Z	(b) Площадь между средним и Z	(c) Площадь за пределами Z
[illegible]	[illegible]	[illegible]
2,92	0,4982	[illegible]
2,93	0,4983	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
3,05	0,4989	0,0011
3,06	0,4989	0,0011
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
3,09	0,4990	0,0010
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
3,29	[illegible]	[illegible]

а	b	c
z	Площадь между средним и z	Площадь за пределами z
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0.0005
[illegible]	[illegible]	0.0004
[illegible]	[illegible]	0.0004
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0,4996	0.0004
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4997	0.0003
[illegible]	0.4997	0.0003
[illegible]	[illegible]	0.0003
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0.0003
[illegible]	[illegible]	0.0003
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4998	0.0002
[illegible]	0.4998	0.0002
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.4999	0.000
[illegible]	0.4999	0.0001
[illegible]	0.4999	[illegible]

Б t-распределение

	Уровень значимости для одностороннего критерия					
	0,10	0,05	0,025	0,01	0,005	0,0005
	Уровень значимости для двустороннего критерия					
Степени свободы, df	0,20	[illegible]	[illegible]	[illegible]	0,01	0,001
1	3,078	6,314	12,706	[illegible]	63,657	[illegible]
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	1,476	2,015	2,571	3,365	[illegible]	[illegible]
6	1,440	1,943	2,447	[illegible]	[illegible]	[illegible]
[illegible]	1,415	1,895	2,365	2,998	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	1,383	1,833	2,262	2,821	[illegible]	[illegible]
10	1,372	[illegible]	2,228	2,764	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	1,771	2,160	2,650	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15	1,341	1,753	2,131	2,602	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	1,333	1,740	2,110	2,567	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	1,729	2,093	2,539	[illegible]	[illegible]
[illegible]	1,325	1,725	2,086	2,528	[illegible]	[illegible]
[illegible]	1,323	1,721	2,080	[illegible]	[illegible]	[illegible]
[illegible]	1,321	1,717	2,074	2,508	[illegible]	[illegible]
[illegible]	[illegible]	1,714	2,069	2,500	[illegible]	[illegible]
[illegible]	1,318	1,711	2,064	2,492	[illegible]	[illegible]
[illegible]	1,316	1,708	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Степени свободы (df)	0.10	0.05	0.025	0.01	0.005	0.0005
	Уровень значимости для двустороннего критерия					
	0.20	0.10	0.05	0.02	0.01	0.001
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

(Уровень значимости для одностороннего критерия — верхняя строка заголовка.)

[illegible] Longman Group [illegible]

В Распределение Хи-квадрат

df	0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	0.020	[illegible]404	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.408	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	[illegible]	[illegible]	[illegible]	[illegible]	1.005	1.424	2.366	3.665	[illegible]	[illegible]	[illegible]	9.837	[illegible]	[illegible]
4	0.297	[illegible]	[illegible]	[illegible]	1.649	[illegible]	[illegible]	4.878	[illegible]	7.779	[illegible]	[illegible]668	13.27[illegible]	[illegible]465
5	0.554	0.75[illegible]	[illegible]	1.610	[illegible]	[illegible]	[illegible]	6.064	[illegible]	9.236	[illegible]	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]	[illegible]204	3.070	3.828	5.348	7.231	8.558	[illegible]	[illegible]	[illegible]	[illegible]	22.457
7	[illegible]239	[illegible]564	[illegible]	[illegible]	3.822	4.671	6.346	8.383	[illegible]	[illegible]	14.067	[illegible]622	[illegible]	24.322
8	1.646	2.032	[illegible]	[illegible]	4.594	5.527	7.344	9.524	[illegible]	[illegible]	15.507	18.168	20.090	[illegible]
9	[illegible]088	2.532	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]656	[illegible]	14.684	16.919	19.679	21.666	[illegible]
10	2.558	3.059	[illegible]	4.865	6.179	7.267	9.342	[illegible]	13.442	15.987	18.307	[illegible]	23.209	29.588
11	[illegible]053	3.609	4.575	[illegible]	6.989	[illegible]	10.341	12.899	14.631	[illegible]	19.675	[illegible]	24.725	[illegible]
12	3.571	4.178	[illegible]	6.304	7.807	9.034	[illegible]	14.011	15.812	18.549	[illegible]026	24.054	[illegible]	[illegible]909
13	4.107	4.765	5.892	7.042	[illegible]	9.926	12.340	[illegible]	[illegible]	[illegible]	[illegible]	25.471	27.688	34.528
14	4.660	[illegible]	[illegible]	7.790	[illegible]	[illegible]	[illegible]	16.222	[illegible]	21.064	23.685	26.873	[illegible]	[illegible]
15	5.229	5.985	[illegible]	8.547	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	24.996	28.259	30.578	[illegible]
16	[illegible]	6.614	[illegible]	9.312	[illegible]	12.624	15.338	[illegible]	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	[illegible]	[illegible]	[illegible]	12.002	[illegible]	16.338	[illegible]	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	[illegible]	9.390	10.865	[illegible]	14.440	17.338	20.601	[illegible]	25.989	28.869	32.346	[illegible]	[illegible]
19	7.633	8.567	[illegible]	[illegible]	[illegible]	15.352	[illegible]	[illegible]	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	[illegible]	12.443	[illegible]	16.266	[illegible]	[illegible]	25.038	28.412	[illegible]	35.020	[illegible]	45.315
21	8.897	9.915	11.591	[illegible]	15.445	[illegible]	20.337	23.858	26.171	29.615	[illegible]	[illegible]	38.932	[illegible]
22	9.542	10.600	12.338	14.041	[illegible]	[illegible]	[illegible]	24.939	27.301	30.813	33.924	37.659	[illegible]	[illegible]
23	10.196	[illegible]	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	[illegible]	11.992	[illegible]	[illegible]	18.062	[illegible]	23.337	27.096	29.553	[illegible]	36.415	40.270	[illegible]	[illegible]
25	11.524	12.697	[illegible]	[illegible]	18.940	[illegible]	24.337	28.172	30.675	34.382	37.652	[illegible]	[illegible]	52.620

df	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
26	12 [illegible]	[illegible]	[illegible]	17,292	19,820	21 [illegible]	25,336	[illegible]	[illegible]	[illegible]	[illegible]	42,[illegible]	[illegible]	[illegible]
27	12,[illegible]	[illegible]	[illegible]	18,114	20,703	22 [illegible]	26,336	[illegible]	[illegible]	[illegible]	[illegible]	44,[illegible]	[illegible]	[illegible]
28	13,565	[illegible]	[illegible]	18,939	21,588	23 [illegible]	27,336	[illegible]	[illegible]	[illegible]	[illegible]	45,[illegible]	[illegible]	[illegible]
29	14,256	[illegible]	[illegible]	19,768	22,475	24 [illegible]	28,336	[illegible]	[illegible]	[illegible]	[illegible]	46,[illegible]	[illegible]	[illegible]
30	14,953	[illegible]	18,[illegible]	20,599	23,364	25,[illegible]	29,336	[illegible]	[illegible]	[illegible]	[illegible]	47,[illegible]	[illegible]	[illegible]

[illegible] Statist[illegible] Tables [illegible] Medic[illegible] (1974) 6-е издание [illegible]

Г F-распределение

p = 0,05

n_2 \ n_1	1	2	3	4	5	6	8	12	24	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	[illegible]	[illegible]	[illegible]
2	[illegible]	9,00	[illegible]	[illegible]	19,30	[illegible]	[illegible]	[illegible]	19,45	19,50
3	[illegible]	9,55	[illegible]	[illegible]	9,01	8,94	8,84	8,74	8,64	[illegible]
4	[illegible]	6,94	6,59	6,39	6,26	6,16	6,04	[illegible]	[illegible]	5,63
5	[illegible]	5,79	[illegible]	5,19	[illegible]	4,95	4,82	4,68	[illegible]	[illegible]
6	[illegible]	[illegible]	4,76	4,53	4,39	4,28	[illegible]	4,00	[illegible]	[illegible]
7	[illegible]	[illegible]	4,35	[illegible]	[illegible]	3,87	3,73	3,57	[illegible]	[illegible]
8	[illegible]	4,46	4,07	3,84	3,69	3,58	3,44	3,28	[illegible]	[illegible]
9	[illegible]	4,26	3,86	3,63	3,48	3,37	[illegible]	3,07	2,90	[illegible]
[illegible]	4,96	4,10	[illegible]	3,48	3,33	3,22	3,07	2,91	[illegible]	2,54
[illegible]	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	[illegible]	[illegible]
[illegible]	4,75	3,88	3,49	3,26	[illegible]	3,00	2,85	2,69	2,50	[illegible]
[illegible]	4,67	3,80	[illegible]	3,18	3,02	2,92	[illegible]	2,60	[illegible]	[illegible]
[illegible]	4,60	[illegible]	3,34	[illegible]	2,96	2,85	2,70	[illegible]	2,35	[illegible]
[illegible]	[illegible]	3,68	[illegible]	3,06	2,90	2,79	2,64	[illegible]	2,29	[illegible]
[illegible]	[illegible]	3,63	[illegible]	3,01	[illegible]	2,74	2,59	[illegible]	[illegible]	[illegible]
[illegible]	4,45	3,59	3,20	2,96	2,81	2,70	2,55	[illegible]	[illegible]	[illegible]
[illegible]	4,41	3,55	3,16	2,93	2,77	2,66	2,51	[illegible]	[illegible]	[illegible]
[illegible]	4,38	3,52	3,13	2,90	2,74	2,63	2,48	[illegible]	[illegible]	[illegible]
[illegible]	4,35	3,49	[illegible]	2,87	[illegible]	2,60	2,45	2,28	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	2,84	2,68	2,57	2,42	[illegible]	2,05	[illegible]
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	[illegible]
23	4,28	[illegible]	[illegible]	2,80	2,64	2,53	2,38	2,20	2,00	[illegible]
24	4,26	3,40	[illegible]	[illegible]	2,62	[illegible]	2,36	[illegible]	[illegible]	[illegible]
[illegible]	4,24	[illegible]	2,99	[illegible]	2,60	2,49	2,34	[illegible]	[illegible]	[illegible]
26	[illegible]	[illegible]	[illegible]	2,74	[illegible]	2,47	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	3,35	[illegible]	[illegible]	2,57	2,46	2,30	[illegible]	[illegible]	[illegible]
[illegible]	4,20	[illegible]	[illegible]	[illegible]	[illegible]	2,44	2,29	[illegible]	[illegible]	[illegible]

… Tables for Biological, Agricultural and Medical …

p = 0.01

			4	5	6	8		
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5981	6106	6234
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	99.36	99.42	99.46
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	26.60
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	13.93
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9.47
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	7 3
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	6.07
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5.28
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4.73
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4.33

	1	2	3	4	5	6	8	12	24	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Д

Использование статистик: идеи для исследовательских проектов

В этом приложении приводятся схемы выполнения четырех исследовательских проектов, каждый из которых требует использования пакетов статистических SPSS или MicroCase для анализа набора данных (как правило, это база данных всеобщего [illegible] GSS за 199[illegible] год). Исследовательские проекты [illegible] [illegible] в разных [illegible] [illegible] [illegible] В первом проекте [illegible] [illegible] статистики [illegible] [illegible] Второй проект связан с [illegible] [illegible] В третьем проекте [illegible] [illegible] В четвертом проекте статистические выводы совмещены с мерами связи (с вариантом для многомерного анализа) его необходимо выполнить после изучения третьей части книги (или четвертой). Большинство проектов могут быть выполнены с использованием любых наборов данных, размещенных на web-сайте данной книги

Проект №1. Описательные статистики

1. Выберите один из наборов [illegible] В этом наборе выберите пять переменных (*примечание ваш преподаватель может определить другое количество переменных*) и используйте команду **Frequencies** (в SPSS) или Univariate (в MicroCase) для получения распределений частот и итоговых статистик (если вы работаете в SPSS нажмите кнопку **Statistics** (Статистики) и включите опцию вычисления среднего, медианы, моды, стандартного отклонения и размаха). Для примеров и инструкций см. демонстрационные примеры 3.1, 3.2 и 4.1. Отметьте всю появляющуюся на экране информацию, относящуюся к процессу вычислений или создайте печатную копию. Список переменных, содержащихся в базе данных GSS за 1998 год, приведен в приложении Ж, а о переменных из других наборах данных можно получить информацию в библиотеке кодов на web-странице данной книги
2. Для каждой переменной постройте столбиковые или линейчатые диаграммы для обобщения всей формы распределения переменной. Примеры и инструкции можно найти в демонстрационном примере 2.3

3. [illegible] распределения частот [illegible] графики и выбор [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] асимметрия [illegible] [illegible] диаграмму и сравни [illegible] 3.7 [illegible] описан общую форму распределения (см. главу 2 [illegible] главу 3) и дисперсию (см. главу 4). Для номинальных переменных [illegible] [illegible] цифровые коды. Например, в переменной class [illegible] [illegible] показывает рабочий [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] между *рабочим классом* и *средним классом*.
4. Далее предлагаются примеры [illegible] итоговых фраз [illegible]

   Для *номинальной переменной* (напр. *пол*) определите моду и предоставьте некоторое [illegible] характера распределения Например

   Большинство респондентов состояли в браке (57 5%), однако нередко встречаются также разведенные (17,4%) и одинокие (21 3%).

   Для *порядковой переменной* (напр. фактор престижа работы) используйте медиану (и возможно, моду) и размах

   Значение медианы уровня престижности составляло 44,3, а размах был ограничен значениями 34 и 87 Наиболее распространенным значением являлось 42 [illegible]

   [illegible] *переменной, принадлежащей к шкале интервалов/отношений* (напр. *плотность населения для* 24 стран), используйте среднее (и возможно, медиану или моду) и стандартное отклонение (и, возможно, размах).

   "В трех странах плотность населения (количество человек на одну квадратную милю) варьировалась в диапазоне от 4,00 до 12,000. Стандартное отклонение составило 1000. В среднем по государству, плотность составляла 351 человек на [illegible] Распределение имеет [illegible] [illegible]

Проект №2. Оценивание

В этом примере вы будете исп[illegible] [illegible] [illegible] характеристик населения С[illegible]А. [illegible] вы будете использовать SPSS или M[illegible] [illegible] формулу 7 3 для нахождения [illegible]

А. Оценивание средних

1. [illegible] имеющие по крайней мере четыре категории [illegible] ющие три переменные, *отличные* от перемен- [illegible] и 7.1. *Примечание: ваш преподаватель может определить другое количество переменных.*

2. SPSS: [illegible] Descriptives (Описательные статистики) для вычис- [illegible] чатную копию. Для вычисления доверительных интервал [illegible] лу 7.2. Повторите процедуру для остальных переменных.

 MicroCase: используйте команду Univariate (Одномерный анализ) и нажмите кноп- [illegible] Statistics [illegible] сти экрана. Доверительные интервалы 99% и 95% будут вычислены автоматичес- [illegible] распределением частот. Зап[illegible] или создайте печатную копию. Повторите процедуру для остальных переменных.

3. Загрузите форму статистического опросного листа (Report Form) для данного проекта с Web-сайта данной книги. Заполните форму для каждой переменной, и в специально отведенном пустом поле напишите заключение об интервале, [illegible] уровне и размере выборки. Вы можете записывать самые простые фразы, словно вы готовите материал для газетной ста[illegible] вы ясно сказать, что оценивается характеристика населения США в целом. Например, заключение может выглядеть следующим образом:

 "На основании случайной выборки 123 [illegible] я рассчитал, что водители в США ездят по межштатным автомагистралям со скоростью 64,46–68,22 миль в час [illegible] доверительный интервал данной оценки составляет 95[illegible]

Б. Оценивание пропорции

4. Выберите три номинальные или порядковые переменные с *двумя или тремя* категориями, отличные от переменных, использовавшихся в упражнении 7.2. *Примечание: ваш преподаватель может определить другое количество переменных.*

5. SPSS: Используйте команду Frequencies (Частоты) для нахождения процента выборки в различных категориях каждой переменной. Смените проценты (помни- [illegible] вычислите доверительные интервалы для одной категории каждой переменной (напр. % женщин для sex), используя формулу 7.3.

 MicroCase: используйте команду Univariate (Одномерный анализ) для нахождения процентов выборки в различных категориях каждой переменной. Смените проценты на пропорции и вычислите доверительные интервалы для одной категории каждой переменной (напр. % женщин для [illegible]

6. Загрузите форму опросного листа для данного проекта с Web-сайта [illegible] Заполните лист для каждой переменной и в специально отведенном пустом поле напишите заключение об интервале, [illegible] уровне и размере выборки. Вы можете записывать самые простые фразы, [illegible] словно вы готовите материал для газетной статьи.

7. *Необязательный пункт:* для каждой вычисл[illegible] соответствующую концепцию из [illegible] второго проекта (Проект 2) [illegible]

Проект №3. Проверка значимости

Для данного проекта используйте БД GSS за 1998 год.

А. t-критерий для двух выборок (глава 9)

1. Выберите [illegible] различные зависимые переменные [illegible] или по порядко[illegible] Выберите [illegible] переменные [illegible] зависимых переменных [illegible] могут [illegible] ные могут измеряться по различным шкалам измерения, а также можно использовать одинаковые независимые переменные для обоих критериев.

2. SPSS: нажмите кнопку Analyze (Анализ), Compare Means (Сравнение средних), а затем Independent Samples T Test (t-критерий для независимых выборок). Присвойте имя вашей зависимой переменной(-ым) в окне Test Variable (Тестовая переменная) [illegible] независимым переменным — в окне Grouping Variable (Переменная группирования). Необходимо также указать значения, используемые для описания групп, которые выделяются для независимой переменной. В качестве примера см. демонстрационный пример 9.1. Запишите результаты выполнения проверки (групповые средние, полученное значение t, значимость, размер выборки) или создайте печатную копию. Повторите процедуру для второй зависимой переменной.

 MicroCase: нажмите кнопку t-Test (t-критерий) в меню Statistics Menu, после чего [illegible] t-Test [illegible] те вашу первую зависимую переменную и нажмите стрелку, чтобы скопировать [illegible] в окно Dependent Variable. [illegible] OK, и вы увидите диаграмму "box and whiskers" (т.е. "ящик с усами" — один из способов визуализации множества данных, относящихся к интервальной шкале). Чтобы выполнить проверку по t-критерию, нажмите кнопку Means (Средние) в окне Statistics (Статистики), расположенном в левой части экрана. Примеры можно найти в демонстрационном примере [illegible] верки (средние по группам, полученное значение t, значимость, размер выборки) или создайте печатную копию. По[illegible] переменной.

3. [illegible] форму [illegible] щенную на web-сайте данной книги.

Б. Дисперсионный анализ (глава 10)

1. Выберите [illegible] различные зависимые переменные, измеряемые по шкале интер[illegible] или порядковые переменные, имеющие три или более знач[illegible]. Выберите независимые переменные, которые, на ваш взгляд, могут быть при[illegible] чиной изменения зависимых переменных и которые имеют 3–5 категорий. Для [illegible] имых переменных можно использовать одинаковые независимые пе[illegible]

2. SPSS [illegible] Analyze (Анализ). Compare Means (Сравнение средних) и One-way Anova [illegible] дисперсионный анализ), после чего на экране по[illegible] One-way Anova [illegible] зависимую переменную в списке пе[illegible] скопировать имя переменной в окно Dependent List (Список зависимых переменных) [illegible] Factor (Фактор). Нажмите [illegible] Statistics (Статистики) и нажмите Descriptive Options [illegible] Continue (Продолжить) и OK [illegible]

MicroCase: В меню Statistics [illegible] команду Anova [illegible] Anova [illegible] Dependent Variable [illegible] Independent Variable [illegible] OK [illegible] "box and whiskers" [illegible] Means [illegible]

3. [illegible] Report Form [illegible]

В. Критерий хи-квадрат (глава 11)

1. Выберите две различные зависимые переменные, измеряемые по любой шкале измерения и принимающие пять или меньше (желательно 2–3) разных значений. Для каждой зависимой переменной выберите независимую переменную, которая, на ваш взгляд, может быть причиной ее изменения. Независимые переменные могут быть измерены по различным шкалам измерений и могут иметь пять или менее (желательно 2–3) категорий. Результаты будет удобнее анализировать, если [illegible] использовать одинаковые независимые переменные.

2. SPSS: Нажмите Analyze (Анализ), Descriptive Statistics (Описательные статистики), а затем Crosstabs (Таблицы сопряженности), после чего на экране появится диалоговое окно Crosstabs. Выделите вашу первую зависимую переменную и скопируйте ее в окно Rows (Строки). Затем выделите независимую переменную и скопируйте ее в окно Columns (Столбцы). Нажмите кнопку Statistics (Статистики), расположенную в нижней части окна, и включите флажок Chi-square (Хи-квадрат). Нажмите Continue (Продолжить) и OK. Запишите результаты или создайте печатную копию. Повторите процедуру для второй зависимой переменной.

MicroCase: В меню Statistics (Статистики) выполните команду Cross-tabulation (Построение таблиц сопряженности), в результате чего на экране появится диалоговое окно Cross-tabulation. Выделите вашу первую зависимую переменную и скопируйте ее в окно Row Variable (Переменная строки), а затем скопируйте независимую переменную в окно Column Variable (Переменная столбца). Нажмите OK. Запишите результаты или создайте печатную копию. Повторите процедуру для второй зависимой переменной.

3. Опишите подробно результаты анализа, используя форму Report Form, доступную на Web-сайте данной книги.

Проект №4. Анализ тесноты и значимости связей

А. Использование таблиц сопряженности

1. В БД GSS за 1998 год выберите один из двух вариантов:
   а. Одна зависимая переменная и три независимые переменные (возможные причины).
   б. Одна независимая переменная и три возможные зависимые переменные (возможные эффекты влияния).

Переменные могут быть измерены по любой шкале, однако должны иметь только несколько 2–5 категорий. Сформулируйте вопросы для исследования или гипотезы о связях между переменными. Убедитесь в том, что предполагаемые причинно-следственные связи имеют смысл и поддаются логическому объяснению.

2. SPSS: Используйте процедуру Crosstabs (Таблицы сопряженности) для построения таблиц сопряженности. (Аналогичные ситуации можно найти в любом демонстрационном примере в конце глав 12, 13 или 14.) Нажмите Analyze (Анализ), Descriptive Statistics (Описательные статистики), Crosstabs (Таблицы сопряженности) и введите зависимую переменную(-ые) в строки, а независимую переменную(-ые) — в столбцы. В диалоговом окне Crosstabs нажмите кнопку Statistics (Статистики) и включите опции хи-квадрат, меры связи фи или V Крамера, а [illegible] гамма [illegible] для каждой таблицы. В диалоговом окне Crosstabs нажмите кнопку Cells (Ячейки) [illegible] процентов по столбцам для каждой [illegible]. Запишите полученные результаты или создайте печатную копию.

MicroCase: Используйте процедуру Cross-tabulation (Таблицы сопряженности) для построения таблиц сопряженности. (Аналогичные ситуации можно найти в любом демонстрационном примере, приводимом в конце глав 12, 13 или 14.) Вве-

[illegible]

3. [illegible] кр[illegible]кого отче-
та о результатах

4. *Необязательный многомерный анализ:* выберите одну из парных связей, полученных вами при выполнении пункта 2, и найдите для данной связи логическую управляющую переменную. Запустите повторно процедуру **Crosstabs** (Таблицы сопряженности) в SPSS или процедуру **Cross-tabulation** в MicroCase для парной связи, но теперь с учетом контроля над третьей переменной. Сравните частные таблицы сопряженности друг с другом и с исходной таблицей сопряженности. Является ли прямой исходная парная связь? Имеются ли факты, свидетельствующие о кажущейся или опосредованной связи? Существует ли интерактивная связь между [illegible] отчет данного проекта.

Б. Использование переменных, измеряемые по шкале интервалов, отношений

1. Из базы данных выберите один из двух вариантов:

   а. Одна зависимая переменная и три независимые переменные (возможные причины).

   б. Одна независимая переменная и три возможные зависимые переменные (возможные эффекты влияния).

   [illegible] также и порядковые [illegible] больше 5 (предпочтительно горез-[illegible]

- **SPSS**: для анализа парных связей используйте процедуру **Regression** [illegible] Scatter[illegible]

   [illegible] **Scatterplot** [illegible] **Regression** [illegible]

3. На web-сайте данной книги можно увидеть примеры составления краткого отчета о полученных результатах.

4. *Необязательный многомерный анализ:* выберите одну из парных связей, полученных вами при выполнении пункта 2, и найдите [illegible] переменную. Запустите процедуру **Regression** [illegible] переменными и проанализируйте результаты. Можно [illegible] переменной [illegible] отчет о результатах анализа и включите его в итоговый отчет данного проекта.

E Введение в программы статистического анализа для ПК (*SPSS for Windows* и *MicroCase*)

Повсеместное использование персональных компьютеров (ПК) в современном [illegible] ...ческих исследовани- [illegible] ПК [illegible]

[illegible] несложно, даже для людей с небольшим опытом работы или вообще без такового. Это приложение поможет вам подготовиться к работе с одной из программ статистического анализа: SPSS for Windows или MicroCase. Уточните, с какой программой вы будете работать в течение учебного курса, и изучайте соответствующие разделы данного приложения. Упражнения, приводимые в конце главы, даются отдельно для каждой программы. В любом случае, если вам уже приходилось использовать компьютерную мышь для команды и щелчка и запуска компьютерной программы, можете смело приступать к изучению программ статистического анализа. Даже если вы совершенно не знакомы с ПК, уже совсем скоро вы поймете, насколько дружелюбны и просты в использовании эти программы. Изучив приложение, вы сможете самостоятельно выполнить упражнения, помещенные в конце многих глав настоящей [illegible]

Прежде чем перейти к изучению, необходимо отметить, что данное приложение является *введением* в SPSS и MicroCase. В нем изложен обзор программы и содержится достаточно информации, необходимой для самостоятельного выполнения заданий, предлагаемых в книге. Однако здесь вы не сможете получить ответы на любые ваши вопросы или найти решение каждой задачи, с которой вы столкнетесь в ходе самостоятельной работы. Поэтому сейчас самое время сказать о том, что обе программы обладают большой и простой в использовании справочной системой. Вам следует тщательно ознакомиться с ее содержимым и научиться пользоваться ею в случае необходимости. Чтобы войти в справочную систему из любой программы, достаточно нажать кнопку Help (Справка), которая находится на панели инструментов в верхней части экрана.

Как SPSS, так и MicroCase [illegible] статистическими пакетами [illegible] статистик по вашей команде. После [illegible] группы наблюдений, вы можете прос[illegible] ...ской информации, не выполняя при этом [illegible]

Откуда столько суеты вокруг изучения [illegible]

[illegible] респондентов, и *единственное*, что вам необходимо узнать — это средний возраст. Как вам известно, чтобы вычислить средний возраст в выборке, необходимо сложить все значения и полученный результат разделить на общее количество наблюдений. Сколько времени вам понадобится, чтобы выполнить сложение 50 двузначных чисел [illegible] микрокалькулятора [illegible] эту операцию вручную)? Если вам удастся вводить по одному значению в секунду (что достаточно быстро) или по 60 значений в минуту, то на ввод всех значений и вычисление среднего возраста вам понадобится около 3–4 минут. Даже если вы будете работать медленно и аккуратно и будете выполнять сложение двух чисел за одну секунду, а в следующую секунду будете проверять математическую операцию, то вся процедура у вас займет около 20 минут. Если бы это была единственная необходимая вам информация, компьютеры и статистические пакеты не позволили бы сэкономить какое-либо время.

И все же, такой пример не очень-то похож на реальный. Обычно исследователю приходится иметь дело не с одной, а с десятками и даже сотнями переменных, а выборки могут содержать в себе сотни и тысячи наблюдений. 50 чисел вы можете сложить друг с другом примерно за 3 или 4 минуты, а как быть с наблюдениями, [illegible] всеми этими числами, не допустив ни одной арифметической ошибки? [illegible] ...аемая ситуация, тем ценнее и полезнее окажутся статистические [illegible]

[illegible] благодаря статистическим пакетам, вы сможете [illegible] вычислений и тратить все время на [illegible]

Е.1. С чего начать — базы данных и компьютерные файлы

[illegible] кет SPSS (или MicroCase дол- [illegible]

[illegible] ления. База данных (БД) представляет собой упорядоченный набор [illegible] формации, в качестве которой могут выступать, к примеру, ответы (отклики). Чтобы [illegible] ПК [illegible] компьютера на дискете или другом носителе информации. В файлах можно сохра- [illegible] обновления. Данные могут храниться в файлах неопределенно долго. В связи с тем, что на выполнение анализа данных может уйти несколько месяцев, возможность хранения информации в БД является дополнительным преимуществом использования ПК.

В примерах для SPSS и MicroCase, приведенных в книге, мы работаем с БД, со- [illegible] Social Survey, GSS), проведенного в 1998 году. В этой БД хранятся ответы совершен[illegible]нолетних американцев на вопросы, касающиеся разнообразных социальных про[illegible]блем. GSS в США проводится ежегодно, начиная с 1972 года, и является основным источником информации для сотен исследовательских проектов, выполняемых про[illegible]ональными социологами. Эта БД представляет собой обширный источник информации об общественном мнении в США и содержит данные практически обо всем — начиная с позиций в отношении проблемы абортов и заканчивая мнениями [illegible]

GSS представляет особую ценность, поскольку респонденты выбирались таким образом, чтобы выборка в целом была репрезентативной для всего населения США. Репрезентативная выборка воспроизводит, в уменьшенном виде, характеристики генеральной совокупности, на основе изучения которой она и была создана (см. главы 6 и 7). Поэтому, если вы будете анализировать БД GSS за 1998 год, вы будете изучать американское общество таким, каким оно было в 1998 году. Все данные [illegible] все связи, которые вам придется анализировать, отражают некоторые из [illegible]олее важных и больных вопросов в жизни американцев.

Полная БД GSS за 1998 год содержит сотни элементов информации (возраст, пол, мнения о таких острых социальных вопросах, как смертная казнь, и т.д.) для почти [illegible] ческими версиями MicroCase или SPSS for Windows [illegible] количество наблюдений и переменных [illegible] ям и уме[illegible]

[illegible]

Е[illegible]

[illegible] поскольку в нем при[illegible] и расшифровка их сути [illegible]

[illegible] проведения исследования приемлемым [illegible] вопросы. NA означает No Answer (Нет о[illegible] однако ответ на него не был получен. DK озна[illegible] [illegible] из анализа. Пропущенные ответы часто встречаются в анализах, однако ввиду их малого количества они не являются помехой.

[illegible] ке процесса анализа данных — за что отвечает статистический пакет (SPSS или [illegible] статистики. Наша задача — применить статистический пакет к БД таким образом, чтобы в конечном итоге получить выходные данные в виде результатов вычислений [illegible] на рис. Е.1.

[illegible]

Рис. Е.1. [illegible]

Статистические пакеты SPSS [illegible] ые и используются для [illegible] ботой с одной лишь GSS. [illegible] тических пакетах. Ра[illegible] Sta[illegible] — каждый из них [illegible]

[illegible]

Е.2. (SPSS) Запуск *SPSS for Windows* и загрузка GSS 1998

[illegible]

Если вы [illegible] будет выведено [illegible]

В этом разделе вы узнали, как [illegible] время открытия файла. [illegible] ить загрузку файла и устан[illegible] некоторые опции вывода информации на экран монитора. Все эти процедуры приведены в таблице Е.1.

Таблица Е.1. Сводка команд (SPSS)

Запустить SPSS [illegible]	[illegible]
Открыть файл данных	[illegible]
Установить опции [illegible] переменных на экран монитора	[illegible] Edit [illegible] Options [illegible] General [illegible] Display Names [illegible] Alphabetical [illegible] и нажмите OK [illegible]

Е.3. (SPSS) Работа с базами данных

[illegible] SPSS Data Editor [illegible] ый столбец представляет переменную или информационный вопрос о проводимом [illegible] шифровку имен переменных можно найти в книге кодов GSS 1998.

[illegible] Utilities [illegible] Variables [illegible] ременной. В правом окне — Variable Information (Информация о переменных) — можно [illegible] менной является *abany*. В окне Variable Information приведен фрагмент вопроса, задаваемого респондентам в ходе исследования (ABORTION IF WOMAN WANTS FOR ANY REASON — может ли женщина сделать аборт по любой причине), а также показаны возможные значения этой переменной (значение 1 соответствует ответу yes — да, а значение 2 соответствует ответу no — нет) и некоторая другая информация.

Такую информацию можно вывести на экран для любой переменной из набора данных. Например, найдите в списке переменную *wrkstat*. Ее можно найти, используя [illegible] Close [illegible] GSS [illegible] тать слева направо) содержатся ответы каждого респондента [illegible]

[illegible] в выделенную верхнюю левую клетку (с рамкой более тем- [illegible] В этой клетке хранится [illegible] [illegible]

тои клавиши.

[illegible] как можно прочитать информацию в окне вывода дан- [illegible] таблице Е.2, и [illegible] GSS 1998.

Таблица Е.2. Сводка команд (SPSS)

[illegible]	[illegible]
[illegible] переменной	[illegible] Utilities [illegible] Variables (Переменные). 2. Прокручивая список имен переменных, найдите имя интересующей вас переменной и выделите его. Информация о переменной появится в правой части окна или 3. См. приложение Ж.

Е.4 (SPSS) Ввод в действие SPSS: вычисление статистик

На данный момент БД выглядит просто как масса чисел, не представляющих для вас никакого значения. Ничего удивительного — ведь вам не придется читать какую-либо информацию на этого экрана. Практически все статистические операции, которые вы будете выполнять, начинаются с выполнения команды Analyze (Анализ) или Statistics (Статистики) в версии 8.0) в строке меню, выбора процедуры и статистик, а потом указании переменной или переменных, которых вы желаете обрабатывать.

В качестве примера давайте попробуем вычислить с помощью *SPSS for Windows* распределение частот переменной *sex*. Распределения частот представляют собой таблицы, в которых показано, сколько раз каждое значение переменной встречается в выборке (см. главу 2). Таким образом, выполнив эту процедуру, мы сможем узнать о количестве мужчин и женщин в выборке GSS 1998.

После загрузки GSS 1998 нажмите в строке меню команду Analyze (Анализ). В [illegible] Descriptive Statistics [illegible] Frequencies [illegible] Frequencies [illegible] рого будут указаны переменные в алфавитном порядке. Будет выделена первая переменная [illegible] списка переменных, или введите с клавиатуры символ s для перехода к приблизительному [illegible]

После выделения интересующей вас [illegible] SPSS [illegible] OK [illegible] Frequencies [illegible]

SPSS [illegible] Output [illegible] SPSS [illegible] Data Editor [illegible] Output [illegible] Data Editor [illegible] Data Editor [illegible] сверните окно Output, нажав значок "–" в верхнем правом углу.

Frequencies (Частоты). Что мы можем сказать на основании данных в этой таблице? [illegible] расположенном справа. Как видно, в выборке имеются данные для 622 мужчин и 765 женщин. В следующих двух столбцах указана информация о процентах, а в последнем столбце справа указана накопленная частота. О последнем столбце мы [illegible]

Один из столбцов процентов имеет обозначение Percent (Процент), а второй — Valid Percent [illegible] обработки отсутствующих значений. Столбец Percent содержит все наблюдения, включая для тех людей, которые вообще не ответили на вопрос (NA), и людей, сообщивших, что они не владеют информацией по данному вопросу (DK). Столбец Valid Percent не содержит пропущенные значения. Так как мы практически всегда стремимся игнорировать пропущенные значения, мы будем работать только со столбцом Valid Percent. [illegible] (пол определялся интервьюерами), и два столбца являются идентичными.

Е.5 (SPSS) Печать и сохранение выходной информации

[illegible] Output [illegible] File [illegible] Print [illegible] SPSS [illegible] File [illegible] Save в памяти компьютера [illegible] (Сохранить), после чего [illegible] OK [illegible] укажите имя [illegible] таблица будет сохранена [illegible]

Е.6. (SPSS) Завершение сеанса работы *SPSS for Windows*

[illegible] результаты своей работы, вы мо- [illegible] File (Файл) и Exit (Вы [illegible] SPSS. В [illegible] вам будет задан вопрос [illegible] желаете ли вы сохранить содержимое окна Output. Вы можете сохранить рас[illegible] NO (Нет). Про [illegible] в котором вы работали [illegible] SPSS.

[illegible] (MicroCase) [illegible] программы [illegible]

Е.7. (MicroCase) Запуск MicroCase и загрузка GSS 1998

Если вы работаете с полной профессиональной версией (Professional) MicroCase, вы будете работать, вероятно, в компьютерной лаборатории, поэтому сможете запустить программу без промедления. Если на своем персональном компьютере вы будете работать с версией программы для студентов (Student), то в первую очередь необходимо установить программное обеспечение. Выполните инструкции, сопровождающие процесс установки программы, и после полного ее завершения можете приступить к изучению данного приложения.

Чтобы начать работу в MicroCase, найдите пиктограмму (или рисунок) на экране вашего монитора с меткой MC, под которой имеется надпись "Student MicroCase". Чтобы переместить на экране монитора стрелку курсора и навести ее на пиктограмму, используйте мышь, и дважды щелкните ее левой кнопкой. После двойного щелчка запустится программа MicroCase.

Через несколько секунд на экране появится окно File and Data (Файл и данные в версии Student) или окно File Management (Управление файлами в версии Professional), в котором вы увидите несколько вариантов выбора. Самой первой операцией, которую необходимо выполнить в MicroCase, является открытие файла данных, поэтому нажмите команду Open File (Открыть файл), расположенную в левом верхнем углу окна. После этого на экран будет выведено окно Open (Открыть) и, вероятно, список файлов данных. Найдите БД General Social Survey 1998, которая может иметь имя GSS98 или подобное имя. Если ваш набор данных хранится на дискете, укажите соответствующий дисковод. Проконсультируйтесь у своего преподавателя, чтобы убедиться, что вы знаете, где найти БД GSS 1998. После этого на экран будет выведено окно File Settings [illegible] некоторая сводная информация о файле данных. Нажмите ОК, и вы вернетесь к [illegible].

Обратите внимание на список команд, [illegible] Help [illegible] ходимо уделить команде Help. Как мы уже говорили, MicroCase имеет обширную встроенную справочную систему, поэтому вам следует ее хорошо изучить самостоятельно.

В этом разделе вы узнали, как запускать M[illegible] в таблице [illegible].

Таблица Е.3. Сводка команд (MicroCase)

Запуск MicroCase	[illegible]
[illegible] файла данных	Нажмите **Open File** (Открыть файл), а затем щелкните на имени файла данных

Е.8. (MicroCase) Работа с базами данных

[illegible] Student [illegible] List Data [illegible], которая находится в нижнем правом углу окна File and Data. Если вы работаете с версией Professional, нажмите команду Data Management [illegible], которая находится в левой части экрана, а затем нажмите List Data. В каждом из случаев вы увидите двумерную сетку в виде столбцов и строк. Таким способом выполняется организация данных в MicroCase. В каждом столбце представлена переменная или статистическая единица БД GSS. Имена переменных указаны вверху столбцов, и вы можете видеть, что первый столбец содержит значения переменной *wrkstat*, во втором столбце содержатся значения переменной *prestg80*, в третьем — *marital* и т.д.

Вы уже знаете, что расшифровку имен переменных GSS 1998 можно найти в книге кодов, которая приводится в приложении Ж. Расшифровать имена переменных можно также и в MicroCase — программа имеет встроенную электронную версию книги кодов GSS 1998. Чтобы воспользоваться ею, нажмите кнопку V, расположенную вверху экрана, после чего на экране появится окно (Переменные).

Это окно состоит из двух частей. В левой части приводится список всех переменных БД, причем первая переменная будет выделена серым цветом. В правой части (Variable Information — информация о переменной) содержится информация о [illegible] окне Variable Information указан вопрос, который задавался респондентам в ходе [illegible] house, or what? — В течение последней недели Вы работали в режиме полных рабочих дней, неполных рабочих дней, были заняты учебой, были заняты домашними работами, были заняты другим делом?), показаны возможные значения этой переменной [illegible] значение 2 соответствует part time — в течение неполных рабочих дней, и т.д.), а также приводится другая информация.

[illegible] положенную вдоль правой границы [illegible] можно также посредством кнопки Search [illegible] того, как вы выделите переменную *marital*, вы може[illegible] респондент [illegible]

[illegible] Что изменяет переменная [illegible] Закройте окно, нажав кнопку Close (Закрыть) [illegible]

[illegible]

[illegible] Last Data [illegible]

[illegible] SPSS 1998

Таблица Е.4. Сводка команд (MicroCase)

[illegible] Data	[illegible]
	Нажимайте стрелки на элементных полосках [illegible] панель правой и нижней [illegible] окна
[illegible]	[illegible] интересующую вас переменную и выделите ее имя. Информация о переменной появится в правой части [illegible]
	[illegible] 3. См. приложение Ж

Е 9. (MicroCase) Использование MicroCase: вычисление статистик

В MicroCase вычисление статистик осуществляется в отдельном окне, которое можно открыть, нажав команду **Statistics Menu** (Меню статистик в версии Student) или команду **Basic Statistics** (Основные статистики в версии Professional). В этом окне представлены многочисленные команды вычисления статистик. Давайте попробуем [illegible] с помощью MicroCase распределение частот для переменной sex. Распределение [illegible] встречается в выборке (см. главу 2). Таким образом, [illegible] **Univariate** [illegible] найти переменную sex, используйте ползунок для прокрутки [illegible] Выделите эту переменную щелчком [illegible] **Primary Variable** [illegible] Результате [illegible] **OK** [illegible] **Summary** [illegible] **Statistics** [illegible] **Output** [illegible]

[illegible] Метки значений male и female [illegible] мужчины и же[illegible] [illegible] Как можно видеть, в выборке [illegible] 765 [illegible] В остальных столбцах [illegible]

Е 10. (MicroCase) Печать выходных данных

[illegible] на печать копию любых полученных статистических данных, щелкнув на пиктограмме **Print** (Печать) в строке меню, или нажав **File** (Файл) и **Print** (Печать).

Е 11. (MicroCase) Завершение сеанса работы в MicroCase

[illegible] **Statistics** [illegible] **Menu** [illegible] **Exit Program** (Выход из программы) [illegible] нетесь к тому окну, в котором вы работали до запуска MicroCase.

Ж Кодировочная таблица для GSS 1998

General Social Survey (GSS) — это опрос общественного мнения, проводимый [illegible] года Национальным советом по изучению общественного мнения (National Opinion Research Council). Версия GSS 1998 размещена на web-[illegible] каждой главы. Наша версия GSS 1998 включает 50 переменных для сформированной [illegible]

[illegible] включает переменные, измеряющие демографические [illegible] фические характеристики респондентов, включая пол, возраст, расовую принадлеж-[illegible] отношение к таким дискуссионным [illegible] как аборты, смертная казнь [illegible]

Многие переменные в наборе данных имеют коды пропущенных значений [illegible] NAP или Not Applicable [illegible] DK — Don't Know [illegible] NA — No Answer — респондент отказался отвечать [illegible]

[illegible] может ли женщина сделать аборт, [illegible]

abany — Она хочет сделать это по любой причине (0. NAP, 8. DK, 9. NA)
1. Да
2. Нет

abhlth — Здоровье женщины находится в опасности
1. Да
2. Нет

age — Возраст респондента (99. NA)
18–89 Фактический возраст

attend — Как часто Вы посещаете церковную службу? (9. DK или NA)
0. Никогда
1. Менее одного раза в год
2. Один или два раза в год
3. Несколько раз в год
4. Примерно один раз в месяц
5. Два-три раза в месяц
6. Практически каждую неделю
7. Каждую неделю
8. Несколько раз в неделю

cappun — Являетесь ли Вы сторонником смертной казни лиц, признанных виновными в совершении убийства? (0. NAP, 8. DK, 9. NA)
1. Да
2. Нет

childs — Сколько у Вас детей? Пожалуйста, укажите всех детей, родившихся живыми (включая детей от предыдущего брака) (9. NA)
0–7 Фактическое количество
8. Восемь и более

class — К какому общественному классу Вы принадлежите (0. NAP, 8. DK, 9. NA)
1. Низший класс
2. Средний класс
3. Рабочий класс
4. Высший класс

cohabit — Согласны ли Вы с тем, что двое людей могут жить вместе, не регистрируя брак? (9. NA)
1. Полностью согласен(-на)
2. Согласен(-на)
3. Отношусь нейтрально
4. Не согласен(-на)
5. Полностью не согласен(-на)

degree — Образование респондента (7. NAP, 8. DK, 9. NA)
0. Незаконченное среднее образование
1. Среднее образование
2. Неполный колледж (сокращенный двухгодичный курс)
3. Бакалавр
4. Высшее образование

educ — Сколько лет Вы потратили в общей сложности на учебу? (97. NAP, 98. DK, 99. NA)

0–20. Фактическое количество лет

eqwlth — Должно ли правительство проводить политику сокращения разницы в уровне доходов между богатыми и бедными [illegible] тому, что правительство ОБЯЗАНО сократить разницу в доходах [illegible]

[illegible] (8. DK, 9. NA, [illegible] NAP)

1. [illegible]
2.
3.
4.
5.
6.
7. [illegible]

fear Есть ли в данном районе такое место (в пределах мили), куда бы Вы [illegible] (0. NAP, 8. DK, 9. NA)

1. Да
2. Нет

fefam Согласны ли Вы с тем, что если в семье мужчина будет зарабатывать [illegible] их обоих? (0. NAP, 8. DK, 9. NA)

1. Полностью согласен(-на)
2. Согласен(-на)
3. Не согласен(-на)
4. Полностью не согласен(-на)

fehelp Согласны ли Вы с тем, что жена должна помогать мужу в его профес- [illegible] (0. NAP, 8. DK, 9. NA)

1. Полностью согласен(-на)
2. Согласен(-на)
3. Не согласен(-на)
4. Полностью не согласен(-на)

grass На Ваш взгляд, нужно ли легализовать употребление марихуаны? (0. NA, 8. DK, 9. NA)

1. Да
2. Нет

gunlaw [illegible] (0. NAP, 8. DK, 9. NA)

1. Да
2. Нет

hapmar Учитывая все факторы, как Вы можете охарактеризовать свой брак? [illegible] (0. NAP, 8. DK, 9. NA)

1. Очень удачный
2. [illegible]
3. [illegible]

happy Учитывая все факторы, как Вы можете охарактеризовать [illegible] (0. NAP, 8. DK, 9. NA)

1. Дела идут [illegible] хорошо
2. Дела идут достаточно хорошо
3. Дела идут не очень хорошо

homosex Что Вы можете сказать о сексуальных отношениях между [illegible] партнерами? Как Вы считаете, [illegible] (5. Other, другое, 0. NAP, 8. DK, 9. NA)

1. Всегда неправильно
2. Практически всегда неправильно
3. Неправильно только в некоторых случаях
4. Нормально

income98 Укажите совокупный доход Вашей семьи в 1998 году [illegible] (98. DK, 99. NA)

1. Менее 1000
2. От 1000 до 2999
3. От 3000 до 3999
4. От 4000 до 4999
5. От 5000 до 5999
6. От 6000 до 6999
7. От 7000 до 7999
8. От 8000 до 9999
9. От 10000 до 12499
10. От 12500 до 14999
11. От 15000 до 17499
12. От 17500 до 19999
13. От 20000 до 22499
14. От 22500 до 24999
15. От 25000 до 29999
16. От 30000 до 34999
17. От 35000 до 39999
18. От 40000 до 49999
19. От 50000 до 59999
20. От 60000 до 74999
21. От 75000 до 89999
22. От 90000 до 109999
23. 110000 и больше

marital В настоящий момент Вы женаты (замужем), вдовец (вдова), разведены [illegible] или никогда не состояли в браке? (9. NA)

1. Женат (замужем)
2. Вдовец (вдова)
3. Разведен(-а)
4. Живу отдельно от жены (мужа)
5. Никогда не состоял(-а) в браке

news Как часто Вы читаете газеты? (0. NAP, 8. DK, 9. NA)

1. Каждый день
2. Несколько раз в неделю
3. Один раз в неделю
4. Менее [illegible] раза в неделю
5. Никогда

[illegible] Сколько [illegible] (97. NAP, 98. DK, 99. NA)

96. Фактически [illegible]

papres80 Престиж профессии отца (0. NAP, DK, NA)
17–86 Фактическое значение

p[illegible] [illegible] партнеров у Вас было в течение последних пяти [illegible]
([illegible] don't know the number — не могу сказать [illegible]сколько, 95. several — несколько, 98. DK, 99. NA, [illegible] NAP)

0. Не было партнеров
1. 1 партнер
2. 2 партнера
3. 3 партнера
4. [illegible] партнера
5. 5–[illegible] партнеров
6. [illegible]–20 партнеров
7. 2[illegible]–[illegible]0 партнеров
8. [illegible] 100 партнеров

p[illegible] Я хочу показать Вам 7-балльную шкалу политических взглядов, от [illegible]либерального до крайне консервативного. Как Вы могли бы охарактеризовать свои политические взгляды? (0. NAP 8. DK 9. NA)

1. Крайне либеральный
2. Либеральный
3. Слегка либеральный
4. Умеренный
5. Слегка консервативный
6. Консервативный
7. Крайне консервативный

po[illegible] Верите ли Вы в жизнь после смерти? (0. NAP 8. DK 9. NA)
1. Да
2. Нет

p[illegible] В обществе бытует множество [illegible] [illegible] [illegible] этого вопроса [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] нормально? ([illegible] NAP 8. DK 9. NA)

1. Всегда не правильно
2. Практически всегда не правильно
3. Не правильно только в некоторых случаях
4. Нормально

ptes96 Вы, вероятно, помните, что в 1996 году Клинтон (Clinton) баллотировался [illegible] президента США от демократической партии [illegible] его [illegible] [illegible] [illegible] [illegible] кандида[illegible]

тов Вы проголосовали (включены только те, кто подтвердил свое участие в голосовании)?
(6. No presidential vote — не проголосовал ни за одного из этих кандидатов. 4. Other — проголосовал за другого кандидата; 0. NAP 8. DK 9. NA)

1. Клинтон
2. Доул
3. Перот

prestg80 Престиж Вашей работы (0. NAP 8. DK 9. NA)
17–86 Фактическое значение

race Расовая принадлежность респондента
1. Белый
2. Темнокожий
3. Другой

racepush Афроамериканцы не должны появляться в тех местах, где их присутствие нежелательно? (0. NAP 8. DK 9. NA)
1. Полностью согласен(на)
2. Практически согласен(на)
3. Практически не согласен(на)
4. Полностью не согласен(на)

region Регион, в котором проводился опрос респондента. (0. NAP 8. DK 9. NA)
1. Новая Англия
2. Среднеатлантическое побережье
3. Центральный северо-восток
4. Центральный северо-запад
5. Южно-Атлантическое побережье
6. Центральный юго-восток
7. Центральный юго-запад
8. Горные штаты
9. Побережье Тихого океана

relig К какой религиозной конфессии Вы относитесь: протестант, католик, исповедую иудаизм, исповедую другую религию, неверующий? (8. DK 9. NA)

1. Протестант
2. Католик
3. Исповедую иудаизм
4. Исповедую другую религию
5. Неверующий

sa[illegible] Как Вы можете охарактеризовать нынешнее финансовое состояние своей семьи в целом: Вы в достаточной мере удовлетворены состоянием, более менее удовлетворены или не удовлетворены? (8. DK 9. NA)

Удовлетворен(-а) в достаточной мере
2. Более-менее удовлетворен(-а)
3. [illegible]

[illegible] Если говорить в целом, насколько Вы удовлетворены своей финансовой [illegible]
(0. NAP, 8. DK, 9. NA)
1. В высшей степени удовлетворен(-а)
2. Более-менее удовлетворен(-а)
3. Несколько неудовлетворен(-а)
4. Крайне неудовлетворен(-а)

[illegible] Пол респондента
1. Мужской
2. Женский

[illegible] Поддержите ли Вы [illegible] в средней школе уроков по сексуальному [illegible]
(0. NAP 3. Depends — в зависимости от обстоятельств), 8. DK, 9. NA)
1. Да
2. Нет

[illegible] Сколько раз (приблизительно) Вы вступали в половую связь в течение [illegible]?
([illegible] NAP, 8. DK, 9. NA)
0. Ни разу
1. Один или два раза
2. Примерно один раз в месяц
3. Два или три раза в месяц
4. Приблизительно каждую неделю
5. Два или три раза в неделю
6. Более трех раз в неделю

[illegible] Численность жителей (в тысячах человек) города/населенного пункта, в котором проживает [illegible]. Данные взяты из отчета о переписи населения в США. Добавьте три [illegible], чтобы закодировать фактические величины

[illegible] Как Вы полагаете, необходимо ли в целях [illegible] наказывать ребенка?
(0. NAP, 8. DK, 9. NA)
1. [illegible] 2. [illegible]
3. Не согласен(-на) 4. Полностью не согласен(-на)

[illegible] Всегда существуют люди, чьи идеи являются шокирующими или опасными для других людей. Имеются в виду те люди, которые [illegible] выступают против любой церкви [illegible]
(0. NAP, 8. DK, 9. NA)

1. Да
2. Нет

[illegible] Как Вы полагаете, имеет ли человек, страдающий тяжелой и неизлечимой болезнью, право на совершение самоубийства?
(0. NAP, 8. DK, 9. NA)
1. Да
2. Нет

[illegible] Говоря в общем, можете ли Вы сказать, что большинству людей Вы смогли бы довериться в чем-либо, или Вы ведете себя с людьми крайне осторожно?
(3. Other — другое, depends [volunteered] — в зависимости от обстоятельств [людям, бескорыстно предлагающим свою помощь])
(0. NAP, 8. DK, 9. NA)
1. Могу доверять большинству людей
2. Крайне осторожен [illegible]

[illegible] [illegible]
([illegible] NAP, [illegible] DK, [illegible] NA)
00–24. Фактическое количество часов

[illegible] Сколько часов в неделю Вы тратите на просмотр телевизионных программ на религиозную тему?
(-1, 999 NAP, 998. DK, 999. NA)
00–24. Фактическое количество часов

[illegible] В течение последней недели Вы: работали полных рабочих дней, неполных рабочих дней, были заняты учебой, были заняты домашними заботами, были заняты другим делом?
1. Работал(-а) в течение полных рабочих дней
2. Работал(-а) в течение неполных рабочих дней
3. Числился(-ась) на работе, но не посещал(-а) ее (в связи с болезнью, отпуском, и т.п.)
4. Безработный(-ая)
5. Пенсионер(-ка)
6. Занят(-а) учебой
7. Занят(-а) домашними заботами
8. [illegible]

[illegible] Смотрели ли Вы в течение [illegible]?
(0. NAP, 8. DK, 9. NA)
1. Да
2. Нет

3 Сводка основных статистических формул и условных обозначений

Глава 2

Пропорция [illegible]

Процентная доля (процентный показатель) [illegible] $\times 100$

Глава 3

Среднее [illegible]

Глава 4

Стандартное отклонение [illegible]

Глава 5

Стандартизованное значение [illegible]

Глава 7

Доверительный интервал выборочного среднего [illegible]

Доверительный интервал выборочной пропорции [illegible]

Глава 8

Среднее [illegible]

Пропорции [illegible]

Глава 9

Среднее [illegible]

Стандартное отклонение распределения выборочных средних [illegible]

Пропорции [illegible]

Стандартное отклонение распределения выборочных пропорций $\sigma_{p-p} = \sqrt{P_u(1-P_u)}\sqrt{(N_1+N_2)/N_1N_2}$.

Обобщённая оценка пропорции генеральной совокупности [illegible]

Глава 10

Общая сумма квадратов $SST = \sum X^2 - N\bar{X}^2$

Межгрупповая сумма квадратов $SSB = \sum N_k(\bar{X}_k - \bar{X})^2$.

Внутригрупповая сумма квадратов $SSW = SST - SSB$

Число степеней свободы для внутригрупповой суммы квадратов $dfw = N - k$

Число степеней свободы для межгрупповой суммы квадратов $dfb = k - 1$.

Внутригрупповая дисперсия $MSW = \frac{SSW}{dfw}$

Межгрупповая дисперсия $MSB = \frac{SSB}{dfb}$

F-отношение $F = \frac{MSB}{MSW}$

Глава 11

[illegible]

Глава 13

Коэффициент фи [illegible]

Коэффициент ... Крамера [illegible]

Коэффициент лямбда [illegible]

Глава 14

Коэффициент гамма $G = \frac{N_s - N_d}{N_s + N_d}$

Коэффициент ро Спирмена $r = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$

Глава 15

Линия регрессии, построенная по методу наименьших квадратов [illegible]

Коэффициент наклона линии регрессии [illegible]

Свободный член [illegible] $- b\bar{X}$

Коэффициент корреляции r Пирсона [illegible]

Глава 17

Коэффициент частной корреляции [illegible]

Уравнение линейной множественной регрессии, полученное по методу наименьших квадратов $Y = a + b_1X_1 + b_2X_2$

Частный коэффициент регрессии при переменной X_1 [illegible]

Частный коэффициент регрессии при переменной X_2 [illegible]

Свободный член уравнения линейной множественной регрессии [illegible]

Бета-коэффициент при переменной X_1 [illegible]

Бета-коэффициент при переменной X_2 [illegible]

Стандартизованное уравнение линейной регрессии, полученное по методу наименьших квадратов [illegible]

Коэффициент множественной детерминации $R^2 = r_{y1}^2 + r_{y2.1}^2(1 - r_{y1}^2)$

Глоссарий символов

a — свободный член (точка, в которой прямая регрессии пересекает ось Y) (глава 15).

AD — среднее отклонение (глава 4).

ANOVA — дисперсионный анализ ДА (глава 4).

b — коэффициент наклона прямой регрессии (глава 15).

b_i — частный коэффициент регрессии при i-ой независимой переменной в уравнении множественной линейной регрессии (глава 17).

b^*_i — стандартизованный частный коэффициент регрессии при i-ой независимой переменной в уравнении множественной линейной регрессии (глава 17).

df — число степеней свободы (глава [illegible]).

f — частота (глава 2).

F — F отношение (глава 10).

[illegible]

f_o — наблюдаемая частота (глава 11).

G — выборочный коэффициент гамма (глава 14).

G_p — частный коэффициент гамма (глава 16).

H_0 — нулевая гипотеза (глава 8);

H_1 — исследовательская или альтернативная гипотеза (глава 8);

IQV — показатель качественной вариации (глава 4);

Md — медиана (глава 3);

Mo — мода (глава 3);

N — количество (число) наблюдений (глава 2);

N_s — число (количество) пар наблюдений, согласованно упорядоченных (глава 14);

N_d — число (количество) пар наблюдений, рассогласованно упорядоченных (глава 14);

% — процентная доля (глава 2);

P — пропорция (глава [illegible]);

P_s — выборочная пропорция (глава 7);

P_u — пропорция в генеральной совокупности (глава 7);

PRE — пропорциональное уменьшение ошибки (глава 13);

Q — интерквартильная широта (глава 4);

r — выборочный коэффициент корреляции Пирсона (глава 15);

r^2 — коэффициент детерминации (глава 15);

R — размах выборки (глава 4);

r_s — выборочный коэффициент ро Спирмена (глава 14);

r — частный коэффициент корреляции (глава 17);

R^2 — множественный коэффициент детерминации (глава 17);

s — выборочное стандартное отклонение (глава 4);

SSB — межгрупповая сумма квадратов (глава 10);

SST — общая сумма квадратов (глава 10);

SSW — внутригрупповая сумма квадратов (глава 10);

s^2 — выборочная дисперсия (глава 4);

t — статистика t-Стьюдента (глава 8);

V — коэффициент V Крамера (глава 13);

X — любая независимая переменная (глава 12);

$\bar{X}$ — выборочное среднее (глава 3);

X_i — любое значение переменной в выборке (глава 3);

Y — любая зависимая переменная (глава 12);

Y' — предсказанное значение для Y (глава 15);

Z scores — стандартизованные значения (Z-значения) (глава 5);

Z — контролируемая (управляющая) переменная (глава 16)

Греческие буквы

α — вероятность ошибки первого рода (глава 8);

β — вероятность ошибки второго рода (глава 8);

γ — коэффициент гамма для генеральной совокупности (глава 14);

λ — коэффициент лямбда (глава 13);

μ — среднее для генеральной совокупности (глава 3);

μ_p — среднее для распределения выборочных пропорций (глава 6);

$\mu_{\bar{X}}$ — среднее для распределения выборочных средних (глава 6);

ρ — коэффициент корреляции Пирсона для генеральной совокупности (глава [illegible]);

ρ_s — ро Спирмена для генеральной совокупности (глава 14);

σ — стандартное отклонение генеральной совокупности (глава 4);

σ_p — стандартное отклонение распределения выборочных пропорций (глава 6);

σ_{p-p} — стандартное отклонение распределения разностей выборочных пропорций (глава 9);

$\sigma_{\bar{X}}$ — стандартное отклонение распределения выборочных средних (глава 6);

$\sigma_{\bar{X}-\bar{X}}$ — стандартное отклонение распределения разностей выборочных средних (глава 9);

σ^2 — выборочная дисперсия (глава 4);

Σ — суммирование по (глава 3);

ϕ — коэффициент фи (глава 13);

χ^2 — статистика хи-квадрат (глава 11);

χ^2_c — статистика хи-квадрат с поправкой Йетса (глава 11).

Ответы на расчетные задачи с нечетными номерами

[illegible] дует попытаться самим решить и интерпретировать задачи, прежде чем обращаться [illegible]

[illegible] или ПК. Я работал с разной степенью точности, которую можно получить, работая с этими устройствами, и вообще-то говоря, не округлял числа до получения конечного результата или до регистрации промежуточной суммы. Я всегда производил округление с точностью до двух знаков (или до второго разряда после десятичной точки или до сотых). Если и вы будете поступать таким же образом, ваши ответы должны быть практически такими же, как и мои. Не удивляйтесь, если они будут немного отличаться, тем более что все отличия должны быть очень несущественными. Если же разница между моими и вашими результатами окажется значительной, выполните повторную проверку всех своих вычислений, чтобы найти возможную допущенную ошибку, или выполните вычисления повторно, но с другой степенью точности.

В заключение позвольте мне сказать вкратце о математических ошибках в данном разделе. Прежде всего, хочу заверить вас в том, что я знаю, насколько важным [illegible] поэтому я не могу дать абсолютной гарантии относительно точности моих расчетов. [illegible] о них, чтобы я в дальнейшем смог внести соответствующие поправки.

Глава 2

2.1. а. Комплекс А: (5/20) × 100 = 25.00%,

Комплекс Б: (10/20) × 100 = 50.00%.

б. Комплекс А: 4:5 = 0.80;

Комплекс Б: 6:10 = 0.60;

в. Комплекс А: (0/20) = 0.00.

Комплекс Б: (1/20) = 0.05.

г. (6/(4 + 6)) = (6/10) = 60.00%.

д. Комплекс А: 8:5 = [illegible]

Комплекс Б: 2:10 = 0.20.

2.3. Удельный показатель ограбления банков = (47/211732) × 100000 = 22.20.

Удельный показатель убийств = (13/211732) × 100000 = 6.14.

Удельный показатель угона автомобилей = (23/211732) × 100000 = 10.86.

Глава 3

3.1 [illegible] *region* [illegible] *support for legalization* [illegible] *number of movies* [illegible] Мода [illegible] сать первокурсников (*N* является четным), поэтому медиана будет равна половине между значениями двух средних наблюдений. Имеем [illegible] старшекурсников (*N* является нечетным), поэтому медиана будет равна значению серединного наблюдения. Чтобы найти среднее для переменных, измеренных по шкале интервалов/отношений, просуммируйте все значения и полученный результат разделите на количество наблюдений.

	Первокурсники	Старшекурсники
Регион места рождения:	Мода = север	Мода = север
[illegible]	Медиана [illegible]	Медиана [illegible] 5
[illegible]	Среднее = [illegible]	Среднее = [illegible]
[illegible]	Среднее = 5.[illegible]	Среднее [illegible]
[illegible]	Медиана [illegible]	Медиана [illegible]
Религия:	Мода = протестант	Мода = протестант и неверующие (каждая по 4 наблюдения)

3.3. [illegible] лого класса медиана равна med [illegible] (средний класс) [illegible] серединное наблюдение находится в этой категории. Средним наблюдением в партии является 26,15. Медианой уровня образования является среднее образование (средняя школа). Модальное семейное положение [illegible] женат (замужем). Среднее количество детей в семье — 2,39.

3.5. Модальная категория семейного положения — married (женат (замужем)). Мода расы — white (белый). Среднее значение возраста [illegible] позиции в отношении к абортам [illegible] 7.

3.7. [illegible] является бимодальной переменной. [illegible] женат (не замужем), с 10 наблюдениями. Медиана [illegible] ренность) равна 3, а средний возраст равен [illegible]

[illegible]

72,50. Таким образом, среднее имеет меньшее значение, и распределение имеет [illegible] асимметрию. Среднее [illegible] равно 78,[illegible]9, а медиана равна 77,5. Распределение имеет очень незначительную положительную асимметрию. К тому же, студенты успешнее справляются с устным тестом, чем с математическим.

[illegible]

Глава 4

[illegible]

[illegible] трудовом участии, однако похожи на мужчин в [illegible] дисперсии двух других переменных. Более детально об этом вопросе [illegible] прочитать в разделе 9.6

4.9.

	$\overline{X}$	s
[illegible]	[illegible]	3,32
Соседи	4,00	4,15

4.11 [illegible]

4.13 а. [illegible]

б. [illegible]

в. [illegible]

г. [illegible]

4.15 [illegible] 12,[illegible]

[illegible] = 13,[illegible]

Глава 5

5.1

X_i	Значение Z	Процент площади выше	Процент площади ниже
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
7	1,00	84,13	15,87
8	-0,67	74,86	25,14
9	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	57,[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

5.3

	Значение Z	Площадь
а.	[illegible]	[illegible]
б.	0,60 и 1,10	10,86%
в.	[illegible]	[illegible]
г.	[illegible]	[illegible]
д.	[illegible]	[illegible]
е.	[illegible]	[illegible]
ж.	[illegible]	[illegible]
з.	[illegible]	[illegible]
и.	[illegible]	[illegible]
к.	[illegible]	[illegible]

5.5

X	Значение Z	Количество студентов выше	Количество студентов ниже
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

X_i	Значение Z	Количество студентов выше	Количество студентов ниже
70	1,06	32	168
82	1,67	10	180
90	3,00	1	199
95	3,83	[illegible]	[illegible]

5 [illegible]

	Значение Z	Площадь
а	[illegible]	[illegible]
б.	1,80	[illegible]
в	[illegible]	[illegible]
г	[illegible]	[illegible]
д	[illegible]	[illegible]
е	[illegible]	[illegible]

5.9

	Значение Z	Площадь
а.	[illegible]	[illegible]
б	[illegible]	[illegible]
в	1,50	0,0668
г	[illegible]	[illegible]
д	[illegible]	[illegible]
е	[illegible]	[illegible]

5.11. Да. Значение строки 110 преобразовывается в значение Z (+2,88). 99,80% площади лежит ниже этого значения, поэтому в данном тесте индивидуум находился выше на [illegible]%

5.13. Для первого случая вероятность составляет 0,0919, а для второго — 0,0655. Первый случай более вероятен

Итоговые упражнения к части I

1 [illegible] каждая из переменных. [illegible] частот и определять соответствующие [illegible] 15 наблюдениями, поэтому медианой будет значение [illegible] наблюдения. Медиана для Strength равна 7, а для Comb[illegible]

[illegible] переменных составляет 9 и 4 соответственно [illegible] имеет смысл использовать стандартное отклонение [illegible] Strength. В [illegible] часто используют для порядковых переменных [illegible] значений. Стандартное отклонение [illegible] переменная [illegible] размах для данной переменной [illegible] QV [illegible]

Для переменной [illegible] среднее равно 1,40, а стандартное отклонение составляет 1,62. Для возраста среднее равно 41,53, а стандартное отклонение составляет 12,54.

3

Количество детей (X):	Среднее = 2,44	s = 2,08
Количество лет, потраченных на учебу (Y)*:	Медиана = 1	Размах = 4
Расовая принадлежность:	Мода = white (белый)	IQV = 0,50
Регулирование рождаемости:	Медиана = 2	Размах = 3
Просмотр телевизионных программ:	Медиана = 3,08	s = 2,04
Религия:	Мода = протестант	IQV = 0,54

Глава 7

7.1. а. [illegible] б. [illegible]
в. 20 ± 0,40; г. 1020 ± 5,41;
д. 7,3 ± 0,23; е. 33 ± 0,80

7.5. а. [illegible] б. [illegible]
в. 6,00 ± 0,37

7.7. а. 178,23 ± 1,97 Оценка: студенты тратят на книги от 176,26 до 180,20 долларов.
б. 1,5 ± 0,04 Оценка: студенты посещают клинику в среднем 1,46–1,54 раза.
в. [illegible] г. [illegible]

7.9. 14 ± 0,07 Оценка. От 7 до 21% населения составляют пары, живущие вместе, но не состоящие в браке.

7.11. а. $P_s = (823/1496) = 0,55$,
Доверительный интервал: 0,55 ± 0,02
53–57% населения согласны с утверждением
б. $P_s = (653/1496) = 0,44$,
Доверительный интервал: 0,44 ± 0,02
в. $P_s = (375/1496) = 0,25$;
Доверительный интервал: 0,25 ± 0,02
г. $P_s = (1023/1496) = 0,68$;
Доверительный интервал: 0,68 ± 0,02
д. $P_s = (800/1496) = 0,54$;
Доверительный интервал: 0,54 ± 0,02.

[illegible]

7.5. [illegible] Оценка от 46 до [illegible] параметр генеральной совокупности P [illegible] может быть как 46%, так и [illegible]

7.7. [illegible] интервал [illegible] 3,98 [illegible]

[illegible]

7.19. а. [illegible]

б. 2,86 ± 0,[illegible]3

в. 1,81 ± 0,09

г. 27 ± 0,03 (около 27% населения – католики).

д. 24 ± 0,03 (около 24% населения никогда не состояли в браке),

е. 53 ± 0,03 (около 49% электората отдали в 1996 году свои голоса в пользу Клинтона).

ж. [illegible]

Глава 8

8.1. а. Z (экспериментальное) = [illegible]

б. Z (экспериментальное) = [illegible]

8.5. Z (экспериментальное) = [illegible]

8.7. а. Z (экспериментальное) = 13,[illegible]

б. Z (экспериментальное) = 25,[illegible]

8.9. t (экспериментальное) = 4,5[illegible]

8.11. Z (экспериментальное) = 3,06

8.13. Z (экспериментальное) = 1,48

8.15. а. Z (экспериментальное) = [illegible]

б. Z (экспериментальное) = [illegible]

в. Z (экспериментальное) = [illegible]

г. Z (экспериментальное) = 18,07

д. Z (экспериментальное) = 2,09

е. Z (экспериментальное) = 53,33

8.17. t (экспериментальное) = -1,[illegible]4

Глава 9

9.1. а. $\sigma = 1{,}39$, Z (экспериментальное) = -0,74;

б. $\sigma = 1{,}60$, Z (экспериментальное) = 2,19

9.3. а. Z (экспериментальное) = 1,70;

б. Z (экспериментальное) = -2,49.

9.5. а. $\sigma = 0{,}08$, Z (экспериментальное) = 11,25;

б. $\sigma =$ [illegible], Z (экспериментальное) = [illegible]3,33

в. $\sigma = 0{,}15$, Z (экспериментальное) = 20,00.

9.7. Это маленькие выборки (совокупное количество менее 100), поэтому в п.4 необходимо произвести вычисления по формулам 9.5 и 9.6.

а. $\sigma = 0{,}12$, t (экспериментальное) = -1,33,

б. $\sigma = 0{,}13$, t (экспериментальное) = 14,85

9.9. t (экспериментальное) = 3,52.

9.11. $P_u = 0{,}45$, $\sigma_p = 0{,}06$, Z (экспериментальное) = 0,67

9.13. а. $P_u = 0{,}46$, $\sigma_p = 0{,}006$, Z (экспериментальное) = 2,17;

б. $P_u = 0{,}80$, $\sigma_p = 0{,}07$, Z (экспериментальное) = 1,43

в. $P_u =$ [illegible]72, $\sigma_p = 0{,}08$, Z (экспериментальное) = 0,75

9.15. а. Z (экспериментальное) = -6,00;

б. Z (экспериментальное) = [illegible]

в. Z (экспериментальное) = -4,33,

г. Z (экспериментальное) = -0,23,

д. Z (экспериментальное) = 5,47;

е. Z (экспериментальное) = -3,56.

Глава 10

10.1. а. Общее среднее равно 12,17, SST = 23[illegible],67, SSB = 173,[illegible]7, а SSW = [illegible] F-отношение = 13,32

б. Общее среднее равно 6,87, SST = 455,73, SSB = 78,53, а SSW = [illegible] F-отношение = 1,25

в. Общее среднее равно 31,65, SST = 8362,55, SSB = 5053,35, а SSW = [illegible] F-отношение = 8,14.

10.3. а. Общее среднее равно 4,39, SST = 86,28, SSB = [illegible], а SSW = [illegible] Экспериментальное значение F-отношения = [illegible] При [illegible] уровне значимости [illegible], равном 0,05 и числах степеней свободы [illegible] критическое значение F-отношения будет равно 3,[illegible] [illegible] отклонить. На принятие решения [illegible] cohabitational (сожительство) [illegible]

[illegible]

10.5. SST = 213,61, SSB = 2,11, а SSW = 211,50. F (экспериментальное) = 0,08.

10.7. SST = 429,48, SSB = 305,42, а SSW = 124,06. F (экспериментальное) = 5,96

Глава 11

11.1 а. 11,[illegible] б. 0,00[illegible] в. 1,52, г. 1,46

11.3 [illegible] нем 0,05, критическое значение критерия хи-квадрат будет равно 3,841. Экспериментальное значение критерия хи-квадрат равно 0,65, поэтому мы не сможем отклонить нулевую гипотезу об отсутствии зависимости между переменными. Между переменными *race (расовая принадлежность)* и *services (уровень обслуживания)* не существует статистически значимой связи.

11.5. С степенью свободы 1 и уровнем альфа, равным 0,05, критическая [illegible] начинается со значения 3,84. Экспериментальное значение критерия хи-квадрат, равное 2,15, не попадает в эту площадь, поэтому нулевую гипотезу отклонить нельзя. Между переменными *unionization (объединение)* и *salary* (з/п) не существует статистически значимой связи.

11.7 Экспериментальное значение критерия хи-квадрат равно 5,13 и является зна[illegible]

11.9 Экспериментальное значение критерия хи-квадрат равно 6,67

11.11 Экспериментальное значение критерия хи-квадрат равно 2,59

11.13 Экспериментальное значение критерия хи-квадрат равно 9,34

11.15 а. Значение критерия хи-квадрат = 25,19

[illegible]

в. Значение критерия хи-квадрат = 5,23

г. Значение критерия хи-квадрат = 28,43

д. Значение критерия хи-квадрат = 14,?7

Итоговые упражнения к части II

1. В эмпирическом анализе необходимо выбрать уместный статистический критерий, который будет использоваться в каждой конкретной ситуации. При [illegible] минимизировать неоднозначность и путаницу в резуль[illegible] правильного решения. В данной [illegible] переменных, с которыми придется работать [illegible]

[illegible] переменной, имеющей только две [illegible] тоды из главы 8 (критерии для одной) [illegible] наличии более двух выборок или [illegible] главы [illegible] не будет работать, если не объединить зн[illegible] *ame Internet* в меньшее количество категорий. [illegible] данной ситуации подходит проверка выборок [illegible] шой выборкой (совокупное N больше 100), поэтому [illegible] разделу 9.2.

Выбрать критерий можно и иначе [illegible] использова[illegible] web-сайте данной книги. Блок-схемы [illegible] можем ответить ДА, мы хотим проверить зна[illegible] к разделу 9.2.

а. $\sigma = 0,14$, Z (экспериментальное) = -35,51

Различие в минутах пользования Internet [illegible] средним) пользуются этой технологией гораздо [illegible]

б. Табличный формат позволяет предположить [illegible] хи-квадрат. Экспериментальное значение критерия [illegible] которое не является значимым на уровне 0,05. [illegible] *(вовлеченность)* и *social class (общественный класс)* [illegible]

Количество партнеров является переменной [illegible] отношений, а образование имеет три категории [illegible] подходящим вариантом является дисп[illegible] ное 0,13, вообще не является значимым [illegible] личного, что подобные величины уровня [illegible] зависимости от уровня образования.

В задаче требуется дать характеристику [illegible] афроамериканцы [illegible] ния дан только по "случайной выборке" [illegible] дура оценки, рассмотренная в главе [illegible] вании уровня альфа, равного 0,05, [illegible]

[illegible] Выборка стати[illegible]

Глава 12

В целях экономии места здесь приводятся ответы только на две задачи

12.1 Эффективность и авторитаризм

	Степень авторитаризма	
Эффективность	Низкая	Высокая
Низкая	37,04%	70,59%
	(10)	(12)
Высокая	62,96%	29,41%
	(17)	(5)
Всего	100,00%	100,00%
	27	17

Условное распределение изменяется, поэтому между переменными имеется связь. Изменение от столбца к столбцу довольно значительное, поэтому связь сильная. Эффективность снижается с ростом степени авторитаризма — рабочие, руководство которых использует диктаторские методы правления, работают малопродуктивно (или, может быть, руководство применяет больше диктаторских методов при невысоком коэффициенте отдачи рабочих), поэтому связь имеет отрицательное направление

12.3 Средний балл отметок в зависимости от привлекательности

	Привлекательность		
Средний балл отметок	Низкая	Средняя	Высокая
Низкий	28,00%	26,67%	33,33%
Средний	40,00%	33,33%	36,56%
[illegible]	32,00%	40,0[illegible]	[illegible]
[illegible]	[illegible]00,00%	100,00%	100,00%

[illegible] распределения изменяются, поэтому между [illegible] переменными [illegible] связь. Изменение от столбца к столбцу довольно [illegible] связь слабая. Можно заметить [illegible] балла [illegible] с возрастанием [illegible] что [illegible] сказать [illegible] слабой), однако характер расположения процентных долей в таблице не подтверждает подобной гипотезы.

Глава 13

13.1 а. $\phi =$ [illegible]

б. $\phi = 0{,}09$, $\lambda = 0{,}00$

в. [illegible]

13.3. $\phi = 0{,}[illegible]7$, $\lambda = 0{,}00$

13.5 [illegible]

13.7 $\phi = 0{,}00$, $\lambda = 0{,}00$

13.9. а. $\phi = 0{,}05$, $\lambda = 0{,}06$

б. $\phi = 0{,}54$, $\lambda = 0{,}33$

в. $\phi = 0{,}39$, $\lambda = 0{,}05$

13.11. V Крамера $= 0{,}[illegible]$, $\lambda = 0{,}[illegible]$

Глава 14

14.1. а. $G = 0{,}71$ б. $G = 0{,}71$ в. $G = -0{,}88$

Эти связи являются достаточно тесными. Уровень [illegible] и доход возрастают с ростом продолжительности пребывания в США [illegible] использования [illegible] процентов [illegible] интерпретировать [illegible] связи. В первой таблице [illegible]% новоприбывших [illegible] плохо владеют английским языком, а 60% "старожилов" очень хорошо знают язык. В этой связи низкие значения одной переменной связаны с низкими значениями другой переменной, и значения возрастают одновременно в двух переменных. Контакты со страной происхождения сокращаются с ростом продолжительности пребывания (-0,88). Большинство "новоприбывших" имеют более высокую степень подобных контактов, в отличие от "старожилов".

14.3 $G =$ [illegible]

14.7. $G = 0{,}22$, Z (экспериментальное) $= 0{,}93$.

14.9. $G = -0{,}[illegible]4$

14.11. $r_s = -0{,}46$; t (экспериментальное) $= -1{,}55$.

14.13. $r = 0{,}70$

Между этнической смешанностью и междоусобицей в данной выборке наблюдается сильная положительная связь. Чем больше различий, тем масштабнее междоусобица.

14.15. а. $G = -0{,}[illegible]4$, б. $G = -0{,}17$

в. [illegible] г. [illegible]

д. $G = 0{,}39$

Глава 15

15.1 Поскольку при нахождении коэффициента наклона позиция [illegible] отказ избирателями в поддержке — является зависимой переменной, Y.

Отказ в поддержке и безработица

$b =$ [illegible]00; $a = 39{,}00$; $Y = (39) + (3)X$;

$r = 0{,}95$, $r^2 = 0{,}90$, t (экспериментальное) $= 4{,}96$.

Отказ в поддержке и уровень образования

$b = 12{,}67$; $a = -94{,}73$; $Y = (-94{,}73) + (12{,}67)X$

$r = 0{,}98$; $r^2 = 0{,}97$; t (экспериментальное) $= 9{,}80$

Отказ в поддержке и проведение негативных кампаний

$b = -0{,}90$, $a = 114{,}01$, $Y = (114{,}01) + (-0{,}90)X$.

$r = -0{,}87$; $r^2 = 0{,}76$; t (экспериментальное) [illegible]

[illegible] коэффициента наклона помните, что *Number of visi* [illegible]

15.5. б.

		Степень роста	Плотность населения	Процент урбанизации
Убийства	$b =$	0,49	0,01	0,008
		4,02	4,72	-0,1
[illegible]		[illegible]	0,58	99,3
		36,07	62,71	8,75
[illegible]		0,04	0,00	0,75
		[illegible]	450,49	-168,40

в. Для показателя роста равного −1 предсказанный показатель самоубийств будет равен 1,53

Для плотности населения, равной 250, предсказанная показатель грабежей будет равен 207,71

Для штата, степень урбанизации которого составляет 50%, предсказанное значение показателя доли угона автомобилей будет равно 249,10

г. Коэффициенты Пирсона (r и r^2)

	Степень роста	Плотность населения	Процент урбанизации
Убийства	0,59(0,35)	0,28(0,08)	0,76(0,57)
[illegible]	[illegible]	[illegible]	0,85(0,[illegible])
[illegible]	[illegible]	[illegible]	0,76(0,[illegible])

15.7 $b = 0{,}05$; $a = 53{,}18$; $r = 0{,}40$; $r^2 = 0{,}16$.

15.9 Престиж и возраст

[illegible]

Работа и количество детей в семье

[illegible]

Просмотр телевизионных программ и количество детей в семье:

[illegible]

Просмотр телевизионных программ и возраст

$a = 2{,}44$; $b = 0{,}02$; $r = 0{,}16$; [illegible]

Количество детей и возраст

[illegible]

Просмотр телевизионных программ и фактор престижа работы

[illegible]

Итоговые упражнения к части III

1. а. Выбор меры связи во многом зависит от шкалы, по которой [illegible] зуемые переменные. Предполагая, что [illegible] *ниями*) и *percent immigrant* (*процент иммигрантов*) [illegible] лов/отношений, для измерения тес[illegible] коэффициент Пирсона. Эта связь [illegible] $r = -0{,}47$ а процент иммигрантов объясняет 22% [illegible] вонарушений для данных десяти городов. Связь имеет отр[illegible] ление, а это означает, что с ростом процента иммигрантов [illegible] правонарушений.

б. [illegible]

в. Подходящей мерой для порядковых переменных [illegible] ется коэффициент ро Спирмена, который для [illegible] [illegible] образования.

г. *Расовая принадлежность* является номинальной переменной [illegible] размерность больше 2×2. Коэффициент V Крамера [illegible] ренную связь между переменными. Процесты по [illegible] чения этой выборки стремятся к исключению [illegible] совой принадлежности в пользу None of the above (Ни одна [illegible]. Второй, более слабый тренд наблюдается для людей, имеющих [illegible] даже при условии существования связи между этими переменными.

Глава 16

16.1 а. $G = [illegible]71$

б. $G = 0{,}78$ (мужчины),

$G = 0{,}65$ (женщины),

в. $G = 0{,}67$ (англо),

$G = 0{,}74$ (латиноамериканцы).

Парная связь сильная и имеет положительное направление [illegible]

[illegible]

G = -0,34 (женщины)

[illegible]

[illegible] Чем дольше продолжительность курса [illegible]

[illegible]

16.5. G = 0,62 (таблица сопряженности)

G = 0,52 (белые)

[illegible]

G = 0,60 (с учетом контроля над расовой принадлежностью).

G = 0,62 (мужчины)

G = 0,62 (женщины).

G = 0,62 (с учетом контроля над полом респондента)

Наблюдается сильная положительная парная связь. Завершение учебной [illegible] щее влияние на белых, чем на темнокожих. Маловероятно, что темнокожие, не окончившие курс обучения, имели работу в течение, как минимум, последнего года. Пол не оказывает влияние на связь. Общая картина говорит о [illegible] прямой связи между завершением курса обучения и работой в течение последнего года, несмотря на существование некоторой взаимосвязи с расовой принадлежностью.

16. Для Right to Abortion (Поддержка легализации абортов) в зависимости от [illegible] дохода и пола

G = 0,14.

0,06 (мужчины)

0,19 (женщины).

[illegible]

G = -0,16 (мужчины)

G = 0,[illegible] (женщины);

Глава 17

17.1. а. Для отказа в поддержке Y и безработи[illegible] связь между переменными X и Y.

б. [illegible] ботицей Z, $r_{yx.z}$ = -0,89. Управляющая переменная [illegible] тарную связь.

в. Отказ в поддержке Y = 70,25 + 2,09 безработицы [illegible] X_2. Для безработицы X_1 = 10 и антирекламы X_2 = [illegible] 58,90.

г. Для безработицы X_1 b^*_1 = 0,66. Для антирекламы X_2 b^*_2 = [illegible]

Безработица оказывает наиболее [illegible] же, чем антиреклама. Обратите внимание [illegible] влияние независимых переменных [illegible]

д. R^2 = 0,[illegible]

17.3. а. Для междоусобиц Y и безработицы X, с учетом контроля [illegible] Z r = 0,79

б. [illegible]

в. Междоусобица Y = (-14,60) + 4,94 безработицы X + [illegible] будет равна 49,2.

г. Для безработицы X_1 b^*_1 = 0,78. Для урбанизации X_2 [illegible]

д. R^2 = 0,65.

17.5. а. Отказ в поддержке Y = 83,80 + (-1,16) демократов X_1 + [illegible]

б. Для X_1 = 0 и X_2 = 5, Y' = 98,25.

в. Z_y = (-1,27)Z_1 + (0,84)Z_2.

R^2 = [illegible]

17.7. а. Z_y = 0,30 имеющих среднее образование Z_1 + (-0,47) [illegible]

б. R^2 = 0,46

Итоговые упражнения к части IV

1. а. Выбор процедуры многомерного анализа во многом [illegible] по которой измеряются переменные и [illegible] измеряемых каждой из переменных. Регрессия [illegible] рывных переменных, измеренных по интер[illegible] данной ситуации мы имеем три переменные [illegible] *ство часов работы и баллы, полученные на вступительных* [illegible] вступительных экзаменах могут измеряться только [illegible] эти значения представлены в широком диапазоне [illegible] ми. Средний балл отметок является зависимой переменной [illegible] сильная отрицательная связь, показывающая, что [illegible] студентов на работе влияет на [illegible]

[illegible] порядка между средним баллом отметок и баллами, по- [illegible] [illegible] [illegible] [illegible] к учебе

[illegible] для переменных количество часов работы [illegible] [illegible]

[illegible]

$[illegible] 0{,}75 Z [illegible] + 0{,}43 Z_2$

[illegible]

Значения бета-коэффициентов показывают, что переменная количества часов работы оказывает более сильное влияние на успеваемость (средний балл отметок), чем баллы, полученные на вступительных экзаменах. Даже для тех студентов, которые были хорошо подготовлены к учебе в колледже и получили высокие баллы на вступительных экзаменах, занятость на работе оказывает негативное влияние на успеваемость (средний балл отметок). Высокое значение коэффициента R^2 означает, что только небольшой процент изменчивости среднего балла отметок (14%) остается необъясненным двумя независимыми переменными.

6. В этой ситуации мы имеем три дихотомические переменные, две порядковые переменные и одну номинальную переменную. С учетом возможной ограниченной вариации, для анализа этих связей будет подходящим метод детали- [illegible] [illegible] окончили учебу за че- [illegible] студентов с высоким уровнем общественной активнос- [illegible] ма равна 0,69, свидетельствуя о сильной связи.

Осуществление контроля над полом студентов позволяет обнаружить некоторое взаимодействие. Гамма для юношей (0, [illegible] ным коэффициентом гамма, а гамма для девушек [illegible] ками, с учетом того, что уровень общест[illegible] для девушек. Это позволяет вы[illegible] что переменную пол необходимо включить в анализ с целью понимания причин различ[illegible] эффектов влияния для юношей и девушек.

Глоссарий терминов

Каждый термин сопровождается кратким описанием с указанием главы, в которой термин был введен.

Alpha error — **альфа-ошибка.** См. *Ошибка первого рода.* Глава 8.

Alpha level (α) — **уровень ошибки альфа.** [illegible] вероятность ошибки первого рода.

(1) При интервальном оценивании — вероятность того, что [illegible] интервал не содержит значение генеральной совокупности. Глава 7.

(2) При [illegible] вероятность [illegible] является верной. Вероятность ошибки первого рода. Глава 8.

Analysis of variance — **дисперсионный анализ.** Статистический метод, подходящий для проверки различий между двумя или более выборочными средними. Глава 10.

ANOVA. См. *Дисперсионный анализ.* Глава 10.

Arrow keys — [illegible] в SPSS — [illegible] используемые для позиционирования курсора на экране монитора. Приложение Б.

Association — **связь.** Взаимосвязь двух (или более) переменных. Говорят, что между двумя переменными существует связь, если распределение одной переменной изменяется в различных категориях или значениях другой переменной. Г[illegible]

Average deviation — **среднее отклонение.** Средняя величина абсолютного отклонения значений вокруг среднего. Глава 4.

Bar chart — **столбчатая диаграмма.** Разновидность графического [illegible] номинальных и порядковых переменных. Глава 2.

Beta error — **бета-ошибка.** См. *Ошибка второго рода.* Глава 8.

Beta-weights (b^*) — **бета-коэффициенты.** Стандартизованные частные коэффициенты регрессии. Глава 17.

Bias — **смещение; систематическая ошибка.** Критерий, используемый [illegible] выборочных статистик для процедур оценивания. [illegible] является несмещенной (не имеет систематической ошибки), когда [illegible] распределения ее выборочных значений равно [illegible] генеральной совокупности. Глава 7.

Bivariate normal distributions — **двумерные нормальные распределения.** [illegible] статистической модели относительно критерия [illegible] коэффициента корреляции r Пирсона о том, что обе [illegible] распределение. Глава 15.

Bivariate table — таблица сопряженности. Таблица, в которой представлена взаи[illegible] [illegible] Гла[illegible] 1[illegible]2.

Cells — ячейки. [illegible] в табли[illegible] [illegible] Глава 11.

Central Limit Theorem — центральная предельная теорема. Теорема, определяющая [illegible] форму распределения выборо[illegible] [illegible]стик при условии большого размера выборки. Глава 6.

χ^2 (critical) — хи-квадрат (критическое). Значение хи-квадрат, являющееся гранич[illegible] [illegible] распределения хи-квадрат. Глава 11.

χ^2 (obtained) — хи-квадрат (экспериментальное). [illegible] мая на шаге 4 пятишагового алгоритма принятия решения. Глава [illegible]

Chi square test — критерий хи-квадрат. Непараметрический критерий [illegible] [illegible] Гла[illegible]

Class intervals — интервалы группирования. [illegible] распреде[illegible] ления частот для тех переменных, которые измеряются по шкале интервалов/ [illegible]

Cluster sampling — кластерный отбор. Метод равновероятностного отбора, который основан на выборе непересекающихся групп, таких как географические области, до того как отбирать элементы из списка генеральной совокупности. Глава [illegible]

Coefficient of determination (r^2) — коэффициент детерминации (r^2). Доля общей из[illegible] [illegible] Глава 15.

Coefficient of multiple determination (R^2) — коэффициент множественной детерминации. Статистика, равная общей доле изменчивости зависимой переменной, [illegible]

Column — столбец. [illegible]

Conditional distribution of Y — Условное распределение переменной Y. [illegible]

Conditional means of Y — Условные средние переменной Y. [illegible]

Confidence interval — доверительный интервал. [illegible]

Confidence level — доверительный уровень, доверительная вероятность. [illegible] уровня значимости (альфа-ошибки) — это вероятность [illegible] параметра генеральной совокупности не по[illegible]падет в доверительный интервал. Глава 7.

Continuous variable — непрерывная переменная. Переменная, единицу измерения которой можно делить до бесконечности. Глава 1.

Cramer's V — Коэффициент V Крамера. Мера связи между двумя переменными [illegible] [illegible] рат. Глава 13.

Critical region (region of rejection) — критическая область (область отклонения статистической гипотезы). [illegible] распределения выборочных [illegible] [illegible]

Cumulative frequency — накопленная (кумулятивная) частота. Столбец в таблице [illegible] общее количество наблюдений в данном интервале и во всех предыдущих интервалах. Глава 2.

Cumulative percentage — накопленный (кумулятивный) процент. Столбец в таблице [illegible] в данном интервале и во всех предыдущих интервалах. Глава 2.

Data — данные. [illegible] информация, представленная в числовом виде. Глава [illegible]

Database — база данных. [illegible] набор связанной информации. Приложение Е.

Data reduction — свертка данных. [illegible] в виде нескольких статистик. Глава 1.

Deciles — децили. Точки разделения распределения значений на доли, кратные десяти. Глава 3.

Dependent variable — зависимая переменная. [illegible] является влиянием одной или нескольких независимых переменных. Главы 1 и 1[illegible]

Descriptive statistics — описательные статистики. Набор статистик, связанных с (1) обобщением распределения одной переменной или (2) измерением связи между двумя или более переменными. Глава 1.

Deviations — отклонения. Расстояния между значениями и средним. Глава 4.

Direct relationship — прямая связь. Многомерная связь, при которой управляющая переменная не оказывает влияние на парную связь. Глава 16.

Discrete variable — дискретная переменная. Переменная, значения которой измерены по одной из дискретных шкал. Основная единица измерения такой переменной не может быть подвергнута разбиению. Глава 1.

Dispersion — разброс. Количество изменчивости или неоднородности в распределении [illegible] Глава 4.

E_1 Для коэффициента лямбда — количество ошибок предсказания, которые могут быть допущены в попытке предсказания категории, к которой будут относиться значения зависимой переменной, игнорируя независимую переменную. Глава 13.

E_2 Для коэффициента лямбда — количество ошибок предсказания, которые могут быть допущены в попытке предсказания категории, к которой будут относиться значения зависимой переменной, учитывая независимую переменную. Глава 13.

Efficiency — эффективность. Степень группировки выборочных [illegible] вокруг среднего их распределения. Глава 7.

EPSEM, Equal Probability of Selection Method — **метод равновероятностного отбора** [illegible] в выборку. Глава 6.

Expected frequency (f_e) — **ожидаемая частота**. Частоты в клетках, которые ожидаются [illegible] при условии отсутствия связи между [illegible].

Explained variation — **объясненная вариация**. [illegible] изменения влиянием переменной X. Глава 15.

Explanation — **объяснение**. См. *Spurious relationship*. Глава 1[illegible]

F ratio — **F-отношение**. В дисперсионном анализе — статистика, лежащая в основе критерия, вычисленная на шаге 4 пятишагового алгоритма принятия решения. Глава 10.

File — **файл**. База данных (или какой-либо другой набор информации), хранящаяся под одним и тем же именем на жестком диске компьютера или на флоппи-диске. Приложение Е.

Five-step model — **пятишаговый алгоритм принятия решения**. Пошаговое руководство выполнения проверки статистических гипотез. Глава 8.

Frequency distribution — **частотное распределение**. Таблица, в которой показано количество наблюдений в каждой категории переменной. Глава 2.

Frequency polygon (line chart) — **многоугольник, или полигон, частот (линейчатая диаграмма)**. Разновидность графического представления частот переменных, измеренных по шкале интервалов/отношений. Глава 2.

Gamma (G) — **коэффициент гамма**. Мера связи порядковых переменных, упорядоченных в виде таблицы. Глава 14.

Histogram — **гистограмма**. Разновидность графического представления частот переменных, измеренных по шкале интервалов/отношений. Глава 2.

Homoscedasticity — **гомоскедастичность** [illegible] переменной X. Глава 15.

Hypothesis — **статистическая гипотеза**. Утверждение относительно взаимосвязи между переменными, принимаемое на основании теории. По сравнению с теорией, гипотеза является более точной, поскольку в ней полностью описываются термины и концепции. Глава 1.

Independence — **независимость**. Нулевая гипотеза при применении критерия хи-квадрат. Две переменные являются независимыми, если классификация на [illegible] на вероятность что [illegible] будет классифицировано в любую определенную категорию второй переменной. Глава 1[illegible].

Independent random samples — **независимые случайные выборки**. [illegible] не влияет на вероятность того, что любое другое определенное наблюдение будет выбрано для остальных выборок. Глава 9.

Independent variable — **независимая переменная**. Переменная, [illegible] предсказания значений, объяснения зависимой переменной. Такая переменная оказывает влияние на зависимую переменную. Главы 1 и 1[illegible].

Index of qualitative variation (IQV) — **коэффициент изменчивости категорий**. Мера разброса для переменных, представленных в таблицах частотного распределения. Глава 4.

Inferential statistics — **статистические выводы**. Область прикладной статистики, которая занимается обобщением и распространением выводов, [illegible] на основании выборки, на всю генеральную совокупность, из которой [illegible] выборка. Глава 1.

Interaction — **взаимодействие**. Многомерная связь, при которой парная связь [illegible] меняется при переходе от категории к категории управляющей переменной. Глава 16.

Interpretation — **интерпретация**. См. *Intervening relationship*. Глава 16.

Interquartile range (Q) — **интерквартильная широта**. Разность между третьей квартилью и первой. Глава 4.

Interval estimate — **интервальная оценка**. См. *Confidence interval*. Глава 7.

Intervening relationship — **опосредованная связь**. Многомерная связь, при которой парная связь становится очень слабой после установления контроля над третьей переменной. Независимая и зависимая переменные связаны [illegible] через управляющую переменную. Глава 16.

Lambda (λ) — **коэффициент лямбда**. Мера связи [illegible] переменных, упорядоченных в таблице сопряженности. Мера связи лямбда основана на подходе PRE (Proportional Reduction of Error [illegible] уменьшение ошибки). Глава 13.

Level of measurement — **шкала измерений**. Математические характеристики переменной, определяемые процессом измерения. Является [illegible] критерием при выборе статистических методов. Глава 1.

Linear relationship — **линейная зависимость**. Зависимость между двумя переменными, при которой точки наблюдений на диаграмме рассеяния [illegible] аппроксимированы прямой линией. Глава 15.

Line chart — **линейчатая диаграмма**. См. *Frequency polygon*.

Marginals — **маргинальные частоты**. Итоговые суммы по строкам [illegible]

Mean [illegible] — **среднее арифметическое** [illegible] Глава [illegible]

Mean square — **средний квадрат**. В дисперсионном анализе [illegible] полученная путем деления внутригрупповой суммы квадратов (Sum of Squares Within, SSW) или межгрупповой суммы квадратов (Sum of Squares Between, SSB) на соответствующее число степеней свободы. Глава [illegible]

Measures of association — **меры связи**. [illegible] связи между переменными. Глава [illegible]

Measures of central tendency — **меры центра распределения**. [illegible] обобщают распределение значений [illegible] репрезентативное значение распределения. Глава 3.

Measures of dispersion — меры разброса. [illegible] распределения [illegible]

Median (Md) — медиана. Точка в распределении [illegible] по обе стороны [illegible] количество [illegible]. Глава 3.

Mean — медиа [illegible] также. Приложение Е.

Microcase [illegible] разработанный для решения задач [illegible]. Приложение Е.

Midpoint — средняя точка. Точка, расположенная посередине между верхней и [illegible] класса. Глава 2.

Mode — мода. [illegible] часто встречающееся значение в выборке или [illegible] категория переменной. Глава 3.

Multiple correlation — множественная корреляция. Многомерный метод [illegible] независимых переменных [illegible]. Глава 17.

Multiple correlation coefficient (R) — коэффициент множественной корреляции. [illegible] корреляции [illegible] независимыми переменными. Глава 17.

Multiple regression — множественная регрессия. Многомерный метод [illegible] эффекты влияния независимых переменных на [illegible] регрессию используется для предсказания значения [illegible] переменной. Глава 17.

Negative association — отрицательная связь. Парная связь, при которой [illegible] переменных происходят в противоположных направлениях. Так, [illegible] переменной происходит [illegible] на одной переменной [illegible] с другой. Глава 12.

Nonparametric test — непараметрический критерий. [illegible] в котором не применяются [illegible] о форме распределения выборочной статистики. Глава 11.

Normal curve — кривая нормального распределения. Теоретическое распределение [illegible] симметрично [illegible] нормально [illegible] стандартное отклонение [illegible] равно 1. Глава 5.

Normal curve table — таблица кривой нормального распределения. Приложение А [illegible] между Z значением и средним стандартизованного распределения. Глава 5.

Null hypothesis (H_0) — нулевая гипотеза. Утверждение об [illegible] при применении статистического критерия. Глава 8.

Observed frequency (f_o) — наблюдаемая частота. [illegible] таблицы сопряженности. Глава 11.

One-tailed test — одностороний критерий. [illegible] «хвост» выборочного распределения. Глава 8.

One-way analysis of variance — однофакторный дисперсионный анализ. Разновидность анализа ДА, в котором изучается эффект влияния одной переменной на другую. Глава 10.

Partial correlation — частная корреляция. Многомерный метод измерения парной связи при условии контроля над другими переменными. Глава 17.

Partial correlation coefficient — коэффициент частной корреляции. Статистика, выражающая тесноту и направление связи между двумя переменными при условии контроля над остальными переменными [illegible] коэффициент частной корреляции при условии контроля над [illegible] переменной. Глава 17.

Partial gamma (G_p) — частный коэффициент гамма. Статистика, выражающая тесноту связи между двумя переменными после удаления влияния третьей переменной. Глава 16.

Partial slopes — частные коэффициенты регрессии. В уравнении множественной регрессии — коэффициенты при независимых переменных, выражающие степень влияния соответствующих переменных на зависимую переменную при условии отсутствия влияния всех остальных независимых переменных. Глава 17.

Partial tables — частные таблицы сопряженности. Таблицы, получаемые в результате контроля над третьей переменной. Глава 16.

Pearson's r (r) — коэффициент корреляции r Пирсона (r). Мера связи между переменными, измеряемыми по шкале интервалов/отношений. Глава 15.

Percentage (%) — процент (процентная доля). Процент вычисляется путем деления количества наблюдений в категории на количество наблюдений во всех категориях и умножения полученного результата на 100. Глава 2.

Percentile — процентиль (перцентиль). Точка в распределении значений, ниже которой попадает определенный процент наблюдений. Глава 3.

Phi (ϕ) — коэффициент фи. Мера связи, основанная на статистике хи-квадрат. Глава 13.

Pie chart — секторная диаграмма. Разновидность графического представления номинальных и порядковых переменных. Глава 2.

Point estimate — точечная оценка. Оценивание параметра генеральной совокупности, при котором получается единственное значение. Глава 7.

Population — генеральная совокупность. Все возможные теоретически наблюдения, представляющих интерес для исследователя. Глава 1.

Positive association — положительная связь. Парная связь, при которой переменные изменяются в одинаковых направлениях. Так, с увеличением значения одной переменной увеличиваются значения другой переменной, причем высокие значения одной переменной связаны с высокими значениями другой. Глава 12.

Proportion — пропорция. Количество наблюдений в категории, деленное на количество наблюдений во всех остальных категориях. Глава 2.

Proportional reduction of error (PRE) — пропорциональное уменьшение ошибки. Подход, лежащий в основе определения и вычисления некоторых мер связи [illegible] количество ошибок, которые можно [illegible] переменной [illegible] переменной. Глава 13.

Quartiles — квартили [illegible] распределение значений на четыре группы, содержащие (по возможности) одинаковое количество [illegible]. Глава 3.

Random samples — случайные выборки. См. *EPSEM*. Глава 6.

Range (R) — размах. [illegible] Глава [illegible]

Ratio — отношение. [illegible] наблюдений в другой категории. Глава 2.

Region of rejection — область отклонения (гипотезы). См. *Critical region*. Глава [illegible]

Regression line — линия регрессии. [illegible] зависимость между двумя переменными. Линия регрессии подгоняется к [illegible] проходит через все условные средние *Y*, либо подходит к ним настолько близко, насколько это возможно. Глава 15.

Replication — репликация. См. *Direct relationship*. Глава 16.

Representative — репрезентативный. Характеристика случайной выборки или выборки, формируемой в соответствии с правилом равновероятностного отбора (EPSEM). Говорят, что выборка является репрезентативной, если она в точности воспроизводит главные характеристики генеральной совокупности, на основе которой она была сформирована. Глава 6.

Research hypothesis — гипотеза исследования (альтернативная гипотеза). Утверждение, противоположное нулевой гипотезе. Глава 8.

Row — строка. Горизонтальное измерение в таблице. Глава 11.

Sample — выборка. Подмножество генеральной совокупности. При статистическом выводе информация анализируется на основе случайной выборки или выборки, сформированной при помощи равновероятностного отбора, а затем обобщается на генеральную совокупность. Глава [illegible]

Sampling distribution — распределение выборочных статистик. Распределение всех возможных выборочных значений определенной статистики. Глава 6.

Scattergram — диаграмма рассеяния. Разновидность графического представления, [illegible] Глава [illegible]

Simple random sample — простая случайная выборка. [illegible] в выборку. Глава 6.

Skew — асимметрия. Мера, позволяющая выражать "скошенность" распределения [illegible] если имеется несколько слишком больших значений, то будет наблюдаться [illegible] значений — отрицательная. Глава 3.

Slope (*b*) — коэффициент наклона. [illegible] на единицу изменения [illegible] Глава 15.

SPSS for Windows — *SPSS for Windows*. [illegible]

Spearman's rho — коэффициент ро Спирмена. [illegible] Глава [illegible]

Specification — спецификация. См. *Interaction*. Глава [illegible]

Spurious relationship — кажущаяся связь. [illegible] связь становится очень слабой после установления контроля [illegible] переменной. Независимая и зависимая переменные [illegible] причинной-следственной связи. Глава 16.

Standard deviation (*s* или *σ*) — стандартное отклонение. Квадратный корень [illegible] мы квадратов отклонений от среднего деленной на *N* [illegible] ся в качестве меры разброса или рассеяния данных; *s* [illegible] стандартным отклонением, а *σ* — стандартное отклонение [illegible] совокупности. Глава 4.

Standard error of the mean — стандартная ошибка среднего. Стандартное отклонение распределения выборочных средних. Глава 6.

Standardized partial slopes (beta-weights) — стандартизованные частные коэффициенты регрессии (бета-коэффициенты). Частный коэффициент регрессии [illegible] при независимой переменной в стандартизованном уравнении регрессии, выражающий степень влияния этой переменной на зависимую переменную при условии отсутствия влияния других независимых переменных. Глава [illegible]

Statistical package (statpack) — статистический пакет. [illegible] программ, написанных для статистической обработки данных. Приложение [illegible]

Statistics — статистика. Набор математических методов для [illegible] данных. Глава 1.

Stratified random sample — стратифицированная (расслоенная) случайная выборка. Случайная выборка, сформированная из генеральной совокупности посредством выбора наблюдений из подсписков или групп пропорционально [illegible] групп в генеральной совокупности. Глава 6.

Sum of squares between (SSB) — межгрупповая сумма квадратов. Сумма квадратов отклонений выборочных средних от общего среднего [illegible] выборки. Глава 10.

Sum of squares within (SSW) — внутригрупповая сумма квадратов. Сумма квадратов отклонений от категориальных (групповых) средних. Глава [illegible]

Systematic random sample — систематическая случайная выборка. Выборка, сформированная посредством случайного отбора первого [illegible] Глава 6.

***t* (critical) — *t* (критическое)**. [illegible] кой области *t*-распределения. Глава 8.

***t* distribution — *t*-распределение, распределение Стьюдента**. [illegible] зуемое для нахождения критической области [illegible] выборочных средних в случаях, когда [illegible] неизвестным. Глава 8.

***t* (obtained) — *t* (экспериментальное)**. [illegible] этого алгоритма принятия решения [illegible] выборочных средних, когда [illegible] Глава 8.

Test statistic — статистика критерия [illegible] **ного алгоритма принятия решения**, ее выборочное значение располагается в распределении данной статистики. Глава 8.

Theory — теория. [illegible] Глава [illegible]

Total variation — общая вариация. [illegible] ее среднего ($\bar{Y}$). Глава 15.

Two-tailed test — двусторонний критерий. Тип критерия для проверки статистических [illegible] выборочных статистик. Глава 8.

Type I error (alpha error) — ошибка первого рода (альфа-ошибка). Вероятность того, что отклоняемая нулевая гипотеза окажется верной. Глава 8.

Type II error (beta error) — ошибка второго рода (бета-ошибка). Вероятность [illegible] Глава 8.

Unexplained variation — необъясненная вариация. Часть [illegible] переменной Y, которая не может быть объяснена переменной X. Глава 15.

Variable — переменная. [illegible] Глава 1.

Variance (s^2 или σ^2) — дисперсия. [illegible] (σ) [illegible] Глава 4.

Y intercept (a) [illegible] Глава 15.

Z (critical) — Z (критическое). [illegible] указывающее на начало критической [illegible] Глава 8.

Z (obtained) — Z (эмпирическое). Статистика критерия, вычисленная в пункте 4 пятишагового алгоритма. Глава 8.

Z scores — Z-значения. [illegible] Глава 5.

Предметный указатель

А

Альтернативная гипотеза 232, 235

Альфа 200, 233

Асимметрия 95

Б

Бета-коэффициент 532

Бета-ошибка 240

В

Вариация 462

необъясненная 465

объясненная 463

полная 462

Вероятностная выборка 179

Вертикальное измерение 328

Взаимосвязь 492

Выборка 179, 180

кластерная 183

простая случайная 180

расслоенная (стратифицированная) 182

репрезентативная (представительная) 180

систематическая 181

Г

Гамма 414

частная 560

Генеральная совокупность 179, 260, 278

Гипотеза 26, 232

альтернативная 232

исследовательская 26

нулевая 232

Гистограмма 72

Гомоскедастичность 456

Горизонтальное измерение 328

График 70

Д

Данные 23

Двусторонний статистический критерий 235

Детализация [illegible] 491

Дециль 99, 100

Диаграмма 73

линейная 74

рассеяния 451

столбчатая 73

Дисперсия 128

необъясненная 465

Дисперсионный анализ 293

однофакторный 307

Доверительный

интервал [illegible]

уровень 206

И

Интервал группирования 60

Интерквартильная ширина 124

Интерпретация 437

статистик [illegible]

Исследование 25

К

Квартиль 99, 100

Концепция независимости 333

Корреляция

множественная 536

нулевого порядка 525

отрицательная [illegible]

положительная 374

частная 525

коэффициент 525

Коэффициент 500

бета 532

детерминации 464

изменчивости качественной [illegible] 42

корреляции 464

множественной детерминации 536

множественной корреляции 536

наклона (b) 456

фи 19

Кривая нормального распределения 154

Критерии

альтернативные [illegible]

двусторонний [illegible] 235

множественных сравнений 325
непараметрический 327
односторонний 235
хи-квадрат 327
Критическая область 233
Круговая диаграмма 79
Крутизна 137

Л

Линейная связь 453
Линейчатая диаграмма 74
Лямбда 397
вычисление 399

М

Маргинальная частота 329
Медиана 98
Мера
ассоциации 20
разброса 120
связи 335, 342, 364
тесноты связи 30
центра распределения 120
центральной тенденции 120
Множественная регрессия 528
Мода 97

Н

Наименьшая значимая разность 326
Номинальная шкала 57
Нормальные распределения, двумерные 466

О

Область отклонения 233
Обозначения 188
Объяснение 494
Объясняющая связь 494
Однородность дисперсии 466
Односторонний критерий 235
Отклонение 126, 197
Отношение 52
Отрезок 456
Оценка
среднего квадрата 297
точечная 198
Ошибка
альфа 240
бета 240
второго рода 240
первого рода 240

П

Параметр 179
Парные статистические связи 367
Переменная 25
X 366
Y 366
дискретная 32
зависимая 26, 366
независимая 26, 366
непрерывная 32
порядковые 414
непрерывные 414
свернутые 414
управляющая 502
Показатель центра распределения 105
Популяция 31
Порядковая шкала 57
Предсказание 454
Проверка
гипотез 225, 234
значимости 225
односторонняя 235
Пропорция 49
Пропорциональное уменьшение ошибки 396
402
Процедура интервального оценивания 202
Процентиль 99, 100
Процентные показатели 49
Прямая регрессия 452
Пятишаговый алгоритм проверки гипотез 231

Р

Разброс 120
Размах 122
Распределение выборочных статистик 184
Регрессия 454
Репликация 493

С

Свертка данных 30
Свободный член 456
Связь 492
интерактивная 492
кажущаяся 492, 494
непредставительная 494
опосредствованная 492
отрицательная 375
положительная 374
причинно-следственная 365
прямая 493

статистической 377
Сокращение объема данных 30, 48
Составление выборки
независимым случайным образом 260
социально-экономический статус 526
спецификацией 497
Среднее арифметическое 100
Средние точки 62
Среднее отклонение 127
Стандартизованное значение
критическое 233
полученное 233
Стандартное отклонение 126, 128
Стандартная ошибка среднего 167
Статистика 23
дескриптивная 390
критерия 233
описательная 30
фи 391
Статистическая зависимость 377
Статистическиезаключения 32
Столбец 328
Строка 328
Суммарная
частота 62
оценка 262
Суммарный процентный показатель 62
Сумма квадратов
внутри категорий 296
между категориями 296
общая 296

Т

Таблица 491
нормального распределения 157
сопряженности 328, 366
уточнение 491
частная 491
Тау-b Кендалла 414
Теория 25
Терминология 188
Третья переменная 489

У

Уровень 52
Условное
среднее Y 455
распределение переменной 366

Ф

фи 391, 392, 393, 394

Ц

Центральная предельная теорема 188

Ч

Частный коэффициент регрессии 529
стандартизованный 532
Частный наклон 529
стандартизованный 532
Частота
наблюдаемая 331
ожидаемая 331
Частотный полигон 74
Частотное распределение 55
категории 56

Ш

Ширина доверительного интервала 208
Шкала
измерений 33
интервальная/относительная 36
номинальная 35
порядковая 35
статистических измерений 33

Э

Эффект множественных сравнений 312

Я

Ячейка 328

Иностранные термины

χ^2 (критическое) 331
χ^2 (экспериментальное) 331
chi square 327
d Соммерса 414
EPSEM 160
post hoc 309
PRE 396
r Пирсона 460, 465
SES 526
SSB 296
SST 296
SSW 296
t-распределение Стьюдента 241, 242
V Крамера 392
Z-значение 157

Научное издание

Хили Джозеф

Статистика

Социологические и маркетинговые исследования

Технические редакторы *А. А. Рудаков, В. Е. Минин*
Корректор *Т. В. Трифонова*
Компьютерная верстка *Г. А. Бушко*
Художник обложки *К. Ф. Радионов*

Лицензия ИД № 05784 от 07.09.01.
ООО «Питер Принт», 194044, Санкт-Петербург, Б. Сампсониевский пр., д. 29а.
ООО «ТИД «ДС», 04011, Киев-53, ул. Артема 25, [illegible]
ООО «ДиаСофтЮП», 196105, Санкт-Петербург, пр. Ю. Гагарина, д. 1, ком. 108.
Подписано в печать 28.12.06. Формат 70×100 [illegible] Усл. п. л. 51,60. Тираж 3000. Заказ 7
Налоговая льгота — общероссийский классификатор продукции ОК 005-93, том 2; 95 3005 — литература учебная.
Отпечатано с готовых диапозитивов в ОАО «Техническая книга»
190005, Санкт-Петербург, Измайловский пр., 29

Лаборатория Статистических Методов

ЛСМ

Высококлассные специалисты лаборатории готовы оказать консультационную и практическую помощь в применении статистических методов при планировании, проведении, и обработке результатов научных, диссертационных, прикладных исследований с использованием как собственного программного обеспечения так и таких известных статистических пакетов как SPSS, Statistics и др.

У нас богатый опыт решения разнообразных прикладных задач с применением современных методов прикладной статистики в таких областях, как:

- маркетинговые исследования и эконометрика;
- клинические исследования;
- математическое моделирование и оптимизация технологических процессов;
- токсикология и микробиология;
- авиастроение;
- пищевая промышленность и сельское хозяйство;
- материаловедение и композиционные материалы;
- конструирование и технология изготовления режущих инструментов;
- а также многих других областях.

Специалистами лаборатории разработан и успешно применяется пакет прикладных програм **ПРИАМ** (Планирование, Регрессия И Анализ Моделей), позволяющий «на потоке» получать устойчивые многофакторные математические модели в разнообразных практических областях. В 1993 г. он занял третье место в конкурсе програмных продуктов в области фундаментальных и прикладных наук. **ПРИАМ** был приобретен многими организациями и успешно используется ими для решения задач, связанных с получением регрессионных моделей и многокритериальной оптимизацией.

У специалистов лаборатории богатый опыт (более 30 лет) решения прикладных, диссертационных и научных задач с применением статистических методов в реальной, системной постановке. Специалистами лаборатории разрабатываются специализированные програмные средства обработки данных. С ними можно познакомиться на сайте **www.diasoft.kiev.ua**. На этом же сайте вы сможете так же познакомится с некоторыми нашими печатными трудами.

Контактная информация:
Россия, тел.: 8-107-812-251-41-94,
Украина, тел.: 8-044-405-14-47; 8-044-531-90-20
E-mail: [illegible]@diasoft.kiev.ua, babich@carnet.kiev.ua

34350

ИЗДАТЕЛЬСКИЙ ДОМ
ПИТЕР®
WWW.PITER.COM

СПЕЦИАЛИСТАМ КНИЖНОГО БИЗНЕСА!

ПРЕДСТАВИТЕЛЬСТВА ИЗДАТЕЛЬСКОГО ДОМА «ПИТЕР»
предлагают эксклюзивный ассортимент компьютерной, медицинской, психологической, экономической и популярной литературы

РОССИЯ

Москва м. «Калужская», ул. Бутлерова, д. 17б, офис 207, 240; тел./факс (095) 777-54-67; e-mail: sales@piter.msk.ru

Санкт-Петербург м. «Выборгская», Б. Сампсониевский пр., д. 29а, тел. (812) 103-73-73, факс (812) 103-73-83; e-mail: sales@piter.com

Воронеж ул. 25 января, д. 4; тел. (0732) 39-61-70; e-mail: piter-vrn@vmail.ru; piterv@comch.ru

Екатеринбург ул. 8 Марта, д. 267б; тел./факс (343) 225-39-94, 225-40-20; e-mail: piter-ural@r66.ru

Нижний Новгород ул. Премудрова, д. 31а; тел. (8312) 58-50-15, 58-50-25; e-mail: piter@infonet.nnov.ru

Новосибирск ул. Немировича-Данченко, д. 104, офис 502; тел./факс (3832) 54-13-09, 47-92-93, 11-27-18, 11-93-18; e-mail: piter-sib@risp.ru

Ростов-на-Дону ул. Калитвинская, д. 17в; тел. (8632) 95-36-31, (8632) 95-36-32; e-mail: jupiter@rost.ru

Самара ул. Новосадовая, д. 4; тел. (8462)37-06-07; e-mail: piter-volga@sama.ru

УКРАИНА

Харьков ул. Суздальские ряды, д. 12, офис 10–11; тел. (057) 751-10-02, (0572) 58-41-45, тел./факс (057) 712-27-05; e-mail: piter@kharkov.piter.com

Киев пр. Красных Казаков, д. 6, корп. 1; тел./факс (044) 490-35-68, 490-35-69; e-mail: office@piter-press.kiev.ua

БЕЛАРУСЬ

Минск ул. Бобруйская, д. 21, офис 3; тел./факс (37517) 226-19-53; e-mail: piter@mail.by

Ищем зарубежных партнеров или посредников, имеющих выход на зарубежный рынок.
Телефон для связи: **(812) 103-73-73**.
E-mail: grigorjan@piter.com

Издательский дом «Питер» приглашает к сотрудничеству авторов.
Обращайтесь по телефонам: **Санкт-Петербург — (812) 327-13-11**,
Москва — (095) 777-54-67.

Заказ книг для вузов и библиотек: (812) 103-73-73.
Специальное предложение — e-mail: kozin@piter.com

Дж. Хили

СТАТИСТИКА

социологические и маркетинговые исследования

6-е издание

Книга профессора социологии Джозефа Хили является уникальной и современной. В ней рассматриваются актуальные вопросы статистической обработки информации. Большое внимание уделено интерпретации и объяснению статистических характеристик, полученных в результате обработки социологических и маркетинговых данных. Содержится много теоретического материала, позволяющего облегчить использование разнообразных методов математической статистики. В конце каждой главы приводятся примеры статистической обработки данных с помощью современных пакетов — SPSS и MicroCase.

Эта книга станет отличным помощником для студентов и преподавателей социологических, психологических и маркетинговых специальностей, а также для социологов-практиков.

ПИТЕР

издательство DiaSoft
www.diasoft.kiev.ua

THOMSON WADSWORTH

Заказ книг:
197198, Санкт-Петербург, а/я 619
тел.: (812)103-73-74, postbook@piter.com
61093, Харьков-93, а/я 9130
тел.: (057)712-27-05, piter@tender.kharkov.ua

ISBN 5-469-00813-4

9 785469 008132

www.piter.com — вся информация о книгах и веб-магазин